

## SEARCH REQUEST FORM

Requestor's Name: \_\_\_\_\_ Serial Number: \_\_\_\_\_  
Date: \_\_\_\_\_ Phone: \_\_\_\_\_ Art Unit: \_\_\_\_\_

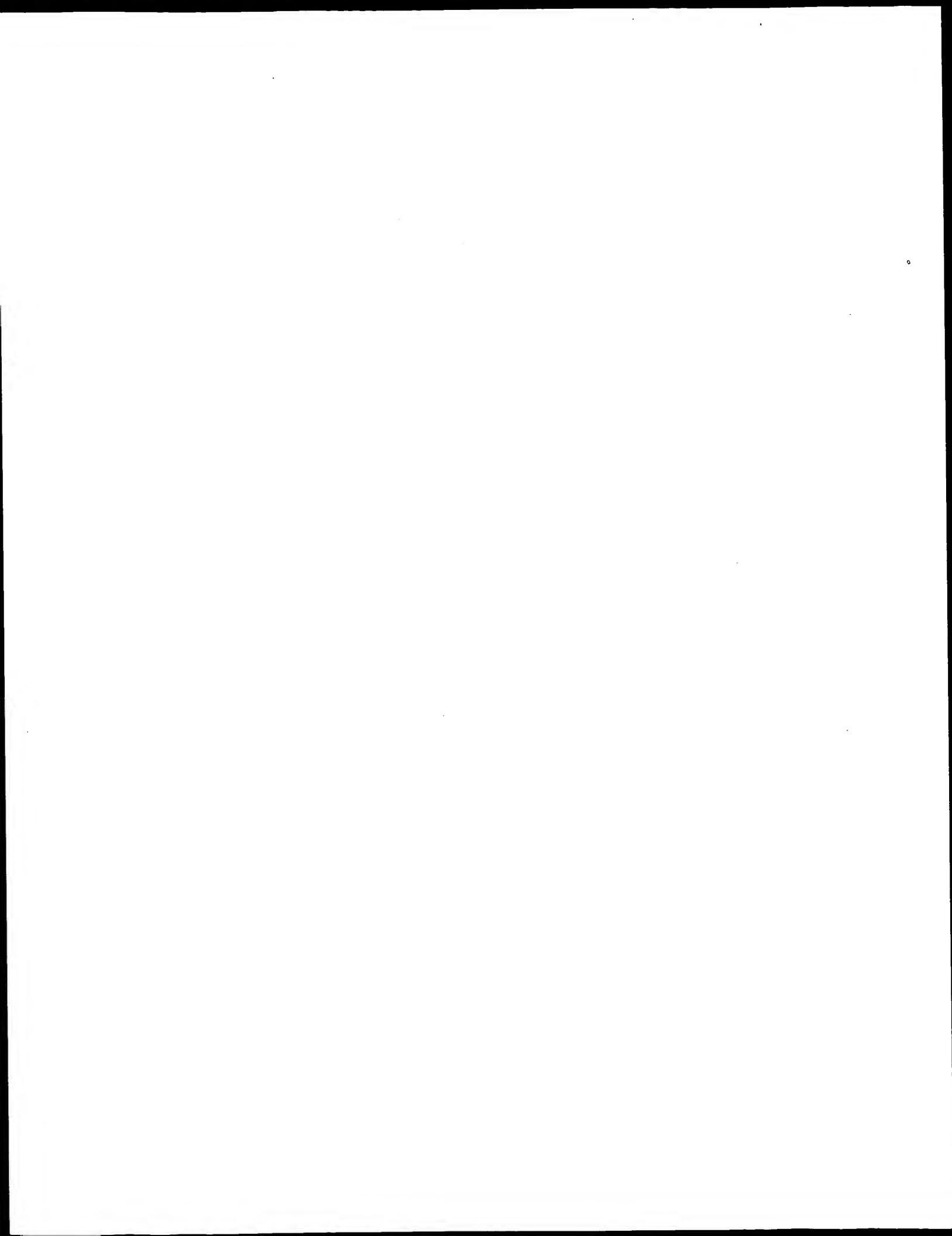
**Search Topic:**

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

**STAFF USE ONLY**

Date completed: 04-16-03  
Searcher: Beverly C 4944  
Terminal time: 23  
Elapsed time: \_\_\_\_\_  
CPU time: \_\_\_\_\_  
Total time: 23  
Number of Searches: \_\_\_\_\_  
Number of Databases: 1

Search Site	Vendors
<input type="checkbox"/> STIC	<input type="checkbox"/> IG Suite
<input type="checkbox"/> CM-1	<input type="checkbox"/> STN
<input type="checkbox"/> Pre-S	<input type="checkbox"/> Dialog
Type of Search	
<input type="checkbox"/> N.A. Sequence	<input type="checkbox"/> APS
<input type="checkbox"/> A.A. Sequence	<input type="checkbox"/> Geninfo
<input type="checkbox"/> Structure	<input type="checkbox"/> SDC
<input type="checkbox"/> Bibliographic	<input checked="" type="checkbox"/> DARC/Questel
	<input checked="" type="checkbox"/> Other C G W



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Run on: April 15, 2003, 13:06:01 ; Search time 3378 Seconds  
(without alignments)  
9175.396 Million cell updates/sec

Title: US-09-905-173-23  
Sequence: 1 atgatacccaagggattaa.....cccttcaagctcttttaa 1065

Scoring table: IDENTITY\_NUC  
Gapext 10.0 , Gapext 1.0

Searched: 2054640 seqs, 11551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: go\_ba: \*  
2: go\_btq: \*  
3: go\_in: \*  
4: go\_om: \*  
5: go\_ov: \*  
6: go\_pat: \*  
7: go\_ph: \*  
8: go\_pl: \*  
9: go\_pr: \*  
10: go\_ro: \*  
11: qb\_sts: \*  
12: qb\_sy: \*  
13: qb\_un: \*  
14: qb\_vl: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hun: \*  
18: em\_in: \*  
19: em\_mlu: \*  
20: em\_on: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pnt: \*  
24: em\_ph: \*  
25: em\_pi: \*  
26: em\_rto: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_hig\_hum: \*  
31: em\_hig\_inv: \*  
32: em\_hig\_other: \*  
33: em\_hig\_mus: \*  
34: em\_hig\_pn: \*  
35: em\_hig\_rod: \*  
36: em\_hig\_man: \*  
37: em\_hig\_vrl: \*  
38: em\_syl: \*  
39: em\_higo\_hum: \*  
40: em\_higo\_mus: \*  
41: em\_higo\_other: \*

Result No.	Score	Query Match	Length	DB ID	Description
1	1065	10.0	1065	6 AR043346	AR043346 Sequence
2	1065	10.0	1065	6 AR078122	AR078122 Sequence
3	1063.4	99.8	12497	1 AR000771	AE000771 Aquifex a
4	145.4	13.7	12075	1 AR013160	AE013160 Thermococcus
5	103.8	9.7	1452	1 AE010265	AE010265 Pyrococcus
6	90.8	8.5	11932	1 AR002508	AE002508 Neisseria
7	90.8	8.5	34980	6 AX044033	AX044033 Sequence
8	89.2	8.4	20334	1 AR001764	AE001764 Thermotoga
9	87.6	8.2	320301	1 NHA62491	AY117255 Arabidopsis
10	86	8.1	22014	1 AR000953	AE000953 Archaeoglobi
11	80.6	7.6	188050	1 AL646072	AL646072 Ralstonia
12	72.2	6.8	1480	1 AF246314	AF246314 Acidithiobacillus
13	68.8	6.5	11265	1 AE012885	AE012885 Chlorobium
14	64.2	6.0	1285	8 AX17255	AY117255 Arabidopsis
15	64.2	6.0	1570	8 AV050832	AY050832 Arabidopsis
16	61	5.7	7218	6 166494	166494 Sequence 14
17	60.2	5.7	17492	1 U67539	U67539 Methanococcus
18	60	5.6	1242	8 NP0278767	AJ228767 Nicotiana
19	56.8	5.6	1556	8 Y09204	Nitrobacillus
20	55.2	5.2	300600	1 AP005369	AP005369 Thermosyn
21	54.4	5.1	13190	1 AB013011	Thermomana
22	46.2	4.3	1085	6 AX431883	AX431883 Sequence
23	45	4.2	30150	1 AP003190	AP003190 Clostridi
24	44.2	4.2	11749	1 AE000964	AE000964 Archaeo
25	44.2	4.2	25770	1 LJU92974	U92974 Lactococcus
26	43.6	4.1	26335	1 CJI1188X5	AL139078 Campylobacter
27	42.2	4.0	1092	6 AR078126	AR078126 Sequence
28	41.6	3.9	119184	1 D90909	D90909 Synechocystis
29	40.8	3.8	10872	1 AE008031	AE008031 Agrobacter
30	40.8	3.8	11490	1 AE000965	AE000965 Agrobacte
31	40.6	3.8	189287	2 AC068362	AC068362 Homo sapi
32	40.6	3.8	200918	9 AC012360	AC012360 Homo sapi
33	40.2	3.8	177585	2 AC094167	AC094167 Rattus no
34	40.2	3.8	245608	2 AC105612	AC105612 Rattus no
35	40	3.8	65785	2 AC10101	AC10101 Rattus no
36	40	3.8	91365	2 AP004776	AP004776 Oryza sat
37	40	3.8	148159	2 AL772390	Mus musculus
38	40	3.8	158050	8 AF480496	AF480496 Oryza sat
39	39.8	3.7	147565	2 AC112868	AC112868 Rattus no
40	39.8	3.7	176744	2 AC128581	AC128581 Rattus no
41	39.8	3.7	201420	2 AC099110	AC099110 Rattus no
42	39	3.7	11080	1 AE006160	AE006160 Pasteurel
43	39	3.6	215296	2 AC093469	AC093469 Mus musculus
44	38.8	3.6	12257	1 AE009793	AE009793 Pyrobaculum
45	38.8	3.6	177014	2 AC022269	AC022269 Homo sapi

## ALIGNMENTS

RESULT 1  
AR043346  
LOCUS AR043346 Sequence 23 from patent US 5814473.

DEFINITION Sequence 23 from patent US 5814473.

ACCESSION AR043346  
VERSION AR043346.1 GI:5964354

KEYWORDS SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1065)

AUTHORS Warren,P.V. and Swanson,R.V.

TITLE Transaminases and aminotransferases

PATENT: US 5814473-A 23-29-SEP-1998;

FEATURES Location/Qualifiers

source	1. .1065 /organism="unknown"	961 GGACITCCAAAAGTGCCTCAGGTAAGCGAACCGGAGAACAACAGRAATTCTG Db 961 GGATTCAAAAGTGCCTCAGGTAAGCGAACCGGAGAACAACAGRAATTCTG 1020 ORIGIN
Query Match	100.0%; Score 1065; DB 6; Length 1065;	QY 1 ATGATACCCAGAGGTTAGGACTTGAACTTGAGCTTACAGAGGGAGTCACCCGCC 60
Best Local Similarity	100.0%; Pred. No. 1. 4e-290; Mismatches 0; Indels 0; Gaps 0;	Db 1 ATGATACCCAGAGGTTAGGACTTGAACTTGAGCTTACAGAGGGAGTCACCCGCC 60
Matches 1065; Conservative		Db 1 ATGATACCCAGAGGTTAGGACTTGAACTTGAGCTTACAGAGGGAGTCACCCGCC 60
QY 1 ATGATACCCAGAGGTTAGGACTTGAACTTGAGCTTACAGAGGGAGTCACCCGCC 60	RESULT 2	QY 1 ATGATACCCAGAGGTTAGGACTTGAACTTGAGCTTACAGAGGGAGTCACCCGCC 60
Db 1 ATGATACCCAGAGGTTAGGACTTGAACTTGAGCTTACAGAGGGAGTCACCCGCC 60	AR078122	AR078122
QY 61 GTGAGGCTTCTCTAACGAAATTCCCTACGACTTTCGGAGATAAACAAAGGGCC 120	LOCUS	Sequence 23 from patent US 5962283.
Db 61 GTGAGGCTTCTCTAACGAAATTCCCTACGACTTTCGGAGATAAACAAAGGGCC 120	DEFINITION	PAT 31-AUG-2000
QY 181 AAAGGGTCTCGGGATTTCGGCTTAGGAGAAATTAGTTCGGTACGGT 240	VERSION	AR078122.1 GI:10004868
Db 181 AAAGGGTCTCGGGATTTCGGCTTAGGAGAAATTAGTTCGGTACGGT 240	KEYWORDS	.
QY 241 TCGGAGGACTCTATACCTACCTCTCATAGCTATAGGTGACTTACACCGTTAC 30	SOURCE	Unknown.
Db 241 TCGGAGGACTCTATACCTACCTCTCATAGCTATAGGTGACTTACACCGTTAC 30	ORGANISM	Unclassified
QY 301 ATGACTGTCCCACCTTCCATGTAGCTAGAGATAAGTGCAGAAGTTCGGAGACCCCTC 360	REFERENCE	1. (bases 1 to 1065)
Db 301 ATGACTGTCCCACCTTCCATGTAGCTAGAGATAAGTGCAGAAGTTCGGAGACCCCTC 360	AUTHORS	Warren, P.V. and Swanson, R.V.
QY 361 GTAAAGGTCACTGGAGAAACTTGATAGACTTAGAAGAAGTATGATAATA 420	TITLE	Transmases and amine transferases
Db 361 GTAAAGGTCACTGGAGAAACTTGATAGACTTAGAAGAAGTATGATAATA 420	JOURNAL	Patent: US 5962283-A 23 OCT-1999;
QY 421 GAGAAAGAACCCGATCTCGGTACTTGTCTACCGAACACCCACGGGAACCTC 480	FEATURES	Location/Qualifiers
Db 421 GAGAAAGAACCCGATCTCGGTACTTGTCTACCGAACACCCACGGGAACCTC 480	source	1. .1065 /organism="unknown"
QY 481 TTTCAGGGAAAGATGTAGGGATAGAAACAGGGGTGTTCTGTATAGCAA 540	BASE COUNT	344 a
Db 481 TTTCAGGGAAAGATGTAGGGATAGAAACAGGGGTGTTCTGTATAGCAA 540	LOCUS	220 c
QY 541 GCCTACTATCATACTCCGGAGAACCTTCGGAGAACCTTCGGAGAACCTAC 600	DEFINITION	100.0%; Score 1065; DB 6; Length 1065;
Db 541 GCCTACTATCATACTCCGGAGAACCTTCGGAGAACCTTCGGAGAACCTAC 600	VERSION	Best Local Similarity 100.0%; Pred. No. 1. 4e-290; Mismatches 0; Indels 0; Gaps 0;
QY 601 GTAGTTGAGGACTTCAAAATCGTATGGCAGTTAGGGTAGGATTATA 660	KEYWORDS	Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 601 GTAGTTGAGGACTTCAAAATCGTATGGCAGTTAGGGTAGGATTATA 660	SOURCE	Unknown.
QY 661 GGAGGGGAATCTACAGGTGAGACTCCCTCAACCTGACCTAC 720	REFERENCE	1. (bases 1 to 1065)
Db 661 GGAGGGGAATCTACAGGTGAGACTCCCTCAACCTGACCTAC 720	AUTHORS	Unclassified
QY 721 CCCCTCAGGTGATGCCAAGTCTCAGGGAGAGATTCTAATGGAAAG 780	TITLE	1. (bases 1 to 1065)
Db 721 CCCCTCAGGTGATGCCAAGTCTCAGGGAGAGATTCTAATGGAAAG 780	JOURNAL	Patent: US 5962283-A 23 OCT-1999;
QY 781 ATACAGGGGTTAACAGAGGATGAGCAAAAGTCTCAGGGAGAGATTCTAATGGAAAG 840	FEATURES	Location/Qualifiers
Db 781 ATACAGGGGTTAACAGAGGATGAGCAAAAGTCTCAGGGAGAGATTCTAATGGAAAG 840	source	1. .1065 /organism="unknown"
QY 901 GAGGGTATCAGGAGACTCTGAAAGGGAGTCTCGTCAGGAGCTACTCATGAA 960	BASE COUNT	344 a
Db 901 GAGGGTATCAGGAGACTCTGAAAGGGAGTCTCGTCAGGAGCTACTCATGAA 960	LOCUS	220 c
QY 961 GGACITCCAAAAGTGCCTCAGGTAAGCGAACCGGAGAACAACAGRAATTCTG Db 961 GGATTCAAAAGTGCCTCAGGTAAGCGAACCGGAGAACAACAGRAATTCTG 1020 ORIGIN	DEFINITION	Sequence 23 from patent US 5962283.
QY 1021 GAAGCACTGGAGGAGATAAATCCCTTCAGCTCAGCTCTCTTAA 1065	VERSION	PAT 31-AUG-2000
Db 1021 GAAGCACTGGAGGAGATAAATCCCTTCAGCTCAGCTCTCTTAA 1065	KEYWORDS	.
QY 421 GAGAAAGAACCCGATCTCGGTACTTGTCTACCGAACACCCACGGGAACCTC 480	SOURCE	Unknown.
Db 421 GAGAAAGAACCCGATCTCGGTACTTGTCTACCGAACACCCACGGGAACCTC 480	ORGANISM	Unclassified
QY 481 GTAAAGGTCACTGGAGAAACTTGATAGACTTAGAAGAAGTATGATAATA 420	REFERENCE	1. (bases 1 to 1065)
Db 481 GTAAAGGTCACTGGAGAAACTTGATAGACTTAGAAGAAGTATGATAATA 420	AUTHORS	Warren, P.V. and Swanson, R.V.
QY 541 GCCTACTATCATACTCCGGAGAACCTTCGGAGACCCGCTCAAAGGGAGATAG 600	TITLE	Transmases and amine transferases
Db 541 GCCTACTATCATACTCCGGAGAACCTTCGGAGACCCGCTCAAAGGGAGATAG 600	JOURNAL	Patent: US 5962283-A 23 OCT-1999;

QY 601 GTAGTGTGAGGACACTTCAAATGGATGGGAGTTAAGGTAGGGATTAA 660  
Db 601 GTAGTGTGAGGACACTTCAAATGGATGGGAGTTAAGGTAGGGATTAA 660  
Db 661 GGGAGGGGAATCTCTGAGAACATTAGAGGAGACCCCTCAAGTGACCTAC 720  
Db 661 GGGAGGGGAATCTCTGAGAACATTAGAGGAGACCCCTCAAGTGACCTAC 720  
Db 721 CCCTCTCAGTGTGATGCCAAAGTCCTCACGGAGGAAGAATCTTAATGAAAG 780  
Db 721 CCCTCTCAGTGTGATGCCAAAGTCCTCACGGAGGAAGAATCTTAATGAAAG 780  
QY 781 ATACAGGAGGTGTACAGCGAGAAGGATGACGACCAATGAGAAATGAGA 840  
Db 781 ATACAGGAGGTGTACAGCGAGAAGGATGACGACCAATGAGAAATGAGA 840  
QY 841 GTTAGGGTTTCCGAGTAAGCTACTTGTGCTTGCAGAGCCTTACCCGCCAC 900  
Db 841 GTTAGGGTTTCCGAGTAAGCTACTTGTGCTTGCAGAGCCTTACCCGCCAC 900  
QY 901 GAGGTATAGGAGCTACTGAAAGGGATGCTCGAGAACATCTTACATGGA 960  
Db 901 GAGGTATAGGAGCTACTGAAAGGGATGCTCGAGAACATCTTACATGGA 960  
QY 961 GGAECTCAAAGTGCCTACGGTAAGCTGGAGAACGGAGAAACACAGTTCTG 1020  
Db 961 GGAECTCAAAGTGCCTACGGTAAGCTGGAGAACGGAGAAACACAGTTCTG 1020  
QY 1021 GAAGCACTGGAGGAGATAAAATCCCTTCAAGCTCTTTAA 1065  
Db 1021 GAAGCACTGGAGGAGATAAAATCCCTTCAAGCTCTTTAA 1065

RESULT<sup>3</sup>

Locus AE000771/C

Definition Aquifex aeolicus section 103 of 109 of the complete genome.

Accession AE000771

Version AE000771.1

Keywords GI:2984286

Organism Aquifex aeolicus.

Reference 1 (bases 1 to 12497)

Authors Decker, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., Overbeek, R., Snead, M.A., Keller, M., AuJay, M., Huber, R., Feldman, R.A., Short, J.M., Olsen, G.J. and Swanson, R.V.

Title The complete genome of the hyperthermophilic bacterium Aquifex aeolicus

Journal Nature 392 (6674), 353-358 (1998)

Medline 98196666

Pubmed 9537320

Reference 2 (bases 1 to 12497)

Title Journal

Comment Submitted (25-JUL-1997) Diversa Corporation, Genomics, San Diego, CA 92121

Putative indicates no similarity to known proteins

Hypothetical indicates similarity to a protein of unknown function.

Location/Qualifiers source

Features

gene

CDS /db\_xref="GI: 2984297" /translation="MESIKETIAFVPEGYGIGTGFIACTTIOVPPDFIAGATA FGINPLISALVATIGLPGVYAVPGRVGPVPVKLGLGEKFKEKLAATYHRGWA VTAGTPIPKVFCWLSGLELEVSRFTLASFIGRPFREFLFALEGHSIQGLF" 387-. 860

gene /gene="aq\_2067" /codon\_start=1 /transl\_table=11 /product="hypothetical protein" /protein\_id="AAC07822.1" /ab\_xref="GI: 2904298" /translation="MLAFGNTPTRSEGVYFGFLYROSEVFPLCRWVTLNSGEPKIKYF IM5FLVIVVSMPILGKESVSYDLSKSGIDLKSKVYLTPMGTIEGEDDEMEVL KKGFEALKEDCNRISTIMKIDYRKGRGRLLQKVKVSQEKGDRDINIAATVEKE" 934-. 1752

CDS /gene="argB" /note="aq\_2068" /codon\_start=1 /transl\_table=11 /product="acetylglutamate kinase" /protein\_id="AAC07812.1" /ab\_xref="GI: 2904288" /translation="MKYGASAMHDEELRESFARDWVLLKYVGINPVWVHGGPQISK TLEFKSTKRVGGMMKTDEEMHVEMWVSGDINKNDIALNIRSGEKYIAYVLSGR DGRLLAKKLKDERYVSELGKLVPEPDIGFGEVGETWNEELIFTISHNPVIAVG VGEEGAYANVADLASELAGEKLTLYLDTKGYLFEGKLSKDKAFELISUSKDKEF LKGVIRGOMIPKVRSALRALERGVKKVHIDERVKHSILLEVFTKEGVGTEITLE" 1765-. 2796

gene /note="aq\_2069" /codon\_start=1 /transl\_table=11 /product="GTP-binding protein" /protein\_id="AAC07816.1" /ab\_xref="GI: 2984292" /translation="MKFKVFRYK\_FWKGGGGDAVAFLREKYRKPKGPGAGDGKGG DVLIWTSKHTLDFRKKHIAQNGEPOKGKNNHGKOGEDLITYVPGTVVKDAQT GEVICOLVKEROKCCTVAGGGGRNARAPPTQNAUTPAEKSGKGEERWILKLLI ADVGFGFPNAGKSULSRTRAKRITAYPTELSPLNPGVMDLWRVLLIADPLI IEDAHHGAGLGEFLRIEHTKFLAHVIDVSDFRREPQQAEEAINREKELSPSKLAQ KPQTIVANKTDALSDRSLSLELEYKEKGYEFYAVSALTGEEGIEELKEGLWKEYEI RDKESEQVT" 2822-. 3997

gene /gene="aq\_2071" /codon\_start=1 /transl\_table=11 /product="hypothetical protein" /protein\_id="AAC07819.1" /ab\_xref="GI: 2984295" /translation="MLQVRVPGEDRIRGFPWVYRSEIASRSRKPKGETIVWVRODV NSKEFQGYINPEVVAIRLISLDFKQMPVPELRKTRIQAYVYKRLYINSNAYRUV VEGLDVWEVRYGEPEEVIWEDDKYIINIPQGKQGFFLDPNMRKVNRLVPEG DRCLDFCHGFGFALMKRAGAGEVYVQDVEISLALKKEKAKNRNLGNDIKWNKAIS DYLRLDKKGEKPFVLLIPPSFKNRVAVDNPNALRQVLPKCYGKLNKSCGYLAIS CSFHTEHILVQASDVRORVYVIAKTFQDLDHPWILOMENTLYLGWVWEVV" 3994-. 5151

CDS /gene="aq\_2073" /codon\_start=1 /transl\_table=11 /product="putative protein" /protein\_id="AAC07818.1" /gene="aq\_2073" /codon\_start=1 /transl\_table=11 /product="putative protein" /protein\_id="AAC07818.1"





Query Match						
	Best Local Similarity	Score	Pred.	No.	Indels	Gaps
	51;	13.78;	145.4;	7.5e-30;	15;	5;
	Conservative			0;	Mismatches	
RBS	QY	16	ATTAAGGAACTTGAACTTACAAAGGGAGTCACGCCCTCCGTCAAGCTTCTCT	49.88;	DB 1;	Length 12075;
misc_feature	Db	4494	ATTAAGGATTTAAATAATGAGTGAATGAAATAGTGAAATAGGGATATGTAAG	44.51		
CDS	QY	76	AACGAATTCCCCTACGACTTCCCGAGGATAAACAAAGGGCTTAGAAA	13.5		
RBS	Db	4434	AACAGGACCCCTTGTACTTCTGTAGAGAAGSAGTAATGAAATATAGGGATATGTAAG	43.75		
RBS	QY	136	AAAGTCTCTTAACAAATACCCAGACCCCAAAGCCAAAGGTTAAAGGGTCTGTG	19.5		
gene	Db	4374	TCTATCCATGAAATATACAGACOCTACGGAGAAATGGAGAGAGATAGC	43.15		
misc_feature	QY	76	TTTGTGGCTTAAGGAGAAATTAGTTCTGGTAAAGGGTACAGTTGTACCC	13.5		
RBS	QY	196	GATTTTGTGGCTTAAGGAGAAATTAGTTCTGGTAAAGGGTACAGTTGTACCC	25.5		
complement	Db	4314	AGATTCGACTCTCACCTCTAAATATTGTAGGAAAGTGTCTGTGAAATAA	42.55		
gene	Db	4254	TACTACCTCTAACAGTATGGACTTTACACCGTTACATACCCGTTCCACC	31.5		
gene	Db	4194	TTGCACTGACGATATGCACTACAGTAAATGCAGGCGAAATACTGTTGATTA	41.95		
complement	Db	316	TTTCCATGACGAGATAGTGGAAAGTCTGGAGACCCCTCTAAGGTCAC	37.5		
gene	Db	4074	TTTGTGGTTTTGTGCAATCCAAACATCCACTGGAAGTATAGTGAAGGAT	40.15		
complement	Db	4134	AGATGAGGACTACTTAAATGTAATGAAAGATGTGAGAAGAGTGAAGATAAGC	40.75		
gene	Db	436	TACCTCGGGTACHTTGTCTTACAGCTAACCCACCGGAACTCTT-TCCAGGGAA	49.3		
complement	Db	553	TTTGTGGTTTTGTGCAATCCAAACATCCACTGGAAGTATAGTGAAGGAT	61.2		
gene	Db	3954	AGATGAGGAGATAAGAACACASGG-GTGTTCCTGTGTTAATAGACGAGCCTACTATCAT	55.2		
complement	Db	4014	ATATACGAGATAATGAAAGAAGGGATGTAAATAGTGTAGAGCTTGTGAA	39.55		
gene	Db	553	TACCTCGGGTACHTTGTCTTACAGCTAACCCACCGGAACTCTT-TCCAGGGAA	66.9		
complement	Db	3894	TTTGTGGTTTTGTGCAATCCAAACATCCACTGGAAGTATAGTGAAGGAT	38.95		
gene	Db	613	ACACTTCAAA--ATCGTATGGCGGTAAAGGGAGATGGTGTGTTAGG	61.2		
complement	Db	3894	ACATGTGCAAAAGTTTGGGATGCGGATTAAGTGGATATGCTCTCAATGGA	38.95		
gene	Db	670	GAATCTGTCAGAAATTCACAAAGGTGAGTCACCCCTCAAGTGACCTACCCCTCAG	72.9		
complement	Db	3834	GAATCTGTCAGAAATTCACCTGTAAATCTCTTATATCTCACTCTTACAG	37.75		
gene	Db	730	GTGAGGGCAAAGTCTCTCACGGAGAGAATCTATGGAAASATACAGGAG	78.9		
complement	Db	3774	AGAATAGCATAGAGTTAA-----AGTGGTTTAAGGAGGTAATTAC	37.21		
gene	Db	790	GTGTGACAGAGGAGAAAGATCTACGACGAATGAGAAATAGAGGGTTGAGTT	84.9		
complement	Db	3720	ATTAAAGGAAAGAGAAGCTGTAAAGGAGTAATTAATTAATGGATAAAAGT	36.61		
gene	Db	850	TTCTGAGTAGGCTACTCTCTCACGGAGAGAATCTATGGCTTACCCGCCACGGTT	9.09		
complement	Db	3660	TATCCCTCTCATGCAATTGTGTCIT--GAAATTGAAATGCAATGATGTCAC	36.04		
gene	Db	910	CAGGCTACTGAAAGGATGTCCTCGTCAAGACGATCTACATGAGAAGACTCAA	9.69		
misc_feature	Db	3603	AAAGACTCTGTGAAAGGGAAATTGTTCTATGAGGGTTGCAA	35.44		
gene	Db	3469	HIS2	45.15		

/note="Pfam match to entry aminotran\_1, Aminotransferase class-1, score -25.9, E-value 1.60E-08"; complement(2255. . 2866);

/gene="HISH"

/note="Best Blastp hit = gi|13913863|sp|066943|HIS5\_AQUAE [Aequifex aeolicus], score 207, E-value 7.00E-53"; complement(3478. . 3482);

/gene="HISB"

/protein\_id="AAM25300\_1"

/db\_xref="GI:20517148", complement(2267. . 2860);

/translation="MIGIDYGAGNLSVEKAIVYGDAKITSDAKEIKWELLIPGVAFDPDIAVRLVKTGMABEIKRHIKEKCPFGICLGHOLIFEEKSFPGKTYFGKLFKGYGEIPAVIKDNFSQFHPKSGERGLKILKNFGEMIKCL"

amidotransferase HISH - Aequifex aeolicus

amidotransferase HISH - Aequifex aeolicus, score 207, E-value 7.00E-53"; complement(2267. . 2860);

/note="Pfam match to entry GATase, Glutamine amidotransferase class-1, score 176.7, E-value 3.80E-49"; complement(2275. . 2279);

/gene="HISA"

/note="Best Blastp hit = gi|387773|sp|033564|HIS7\_RHOSH IMDAOLEGLYCEROL-PHOSPHATE DEHYDRATASE (IGPD) q12398791|emb|CAK60712.1 (X87256) imidazoleglycerol-phosphate dehydratase [Rhodobacter sphaeroides], score 223, E-value 2.00E-57"; complement(2885. . 3469);

/codon\_start=1

/transl\_table=11

/product="Imidazoleglycerol-phosphate dehydratase" /protein\_id="AM25301\_1"

/db\_xref="GI:20517149", complement(2845. . 3379);

/translation="MRKAEKKRKAETDTYVLENIDKGSYDKITKGAGFFDHMSLFA KHHDDIKVIAKGDLVEDPDTHTVLDVGLTALKASGDKKKIKRFTFYVMDEAL VRVSUDISGRPLYIYDPLKAERKVGNEFEETEEFRFRAFYNGQTITHVELLHGNT HIRASFKAIGKALDEALKIDERIDGIPSITKL"

/note="Pfam match to entry IGPD, imidazoleglycerol-phosphate dehydratase, score 293.4, E-value 6.20E-100"; complement(3466. . 4534)

/gene="HIS2"

/note="Best Blastp hit = gi|17450424|pir|1E70478 histidinol-phosphate aminotransferase - Aequifex aeolicus gi|2884289|gb|AAC07813.1 (AE000771) histidinol-phosphate aminotransferase [Aequifex aeolicus], score 230, E-value 2.00E-59"; complement(3466. . 4521);

/codon\_start=1

/transl\_table=11

/product="Histidinol-phosphate aminotransferase" /aminotransferase" /protein\_id="AM25302\_1"

/translation="MENLILREIEIKGFKNEYEVENVYKYMMDANETPFELPEVMKNTGDIKSVTHWNPDPKAKEELARLYCSYTPKNTFVGMSDEIHLIMLAFTVKGDTGVSVIDBEDILITTERKARTVIVDEAYFEEYKGTLVPIYDFRNTIVLRLSKRGIA GLRVGVALSNETVKVLNLYKSPYNLSLSORIALEVLGSLVLRKRVNLTINERKLV BLINKLINGKIPSHANFVLCKFENANDVKEVRLVERGLVLRVNFSNVKGLEGTRITVS SSDKNDYLNAURELIS" complement(3469. . 4515);

/gene="HIS2"

QY 970 AAGTGCCTCAGGTAACCGCTAGGAACCGAGAACACAAGTTCTGGAGCACIG 1029  
 | || | || | || | || | || | || | || | || | || | || | || | || | || |  
 Db 3543 GGGACTTAAGATAACTGTGTCCTCTGCAATGATTAAACGCTTA 3484  
 QY 1030 GAGGAGA 1036  
 |||||  
 Db 3483 AGGGAGA 3477

RESULT 5  
 LOCUS AE010265  
 DEFINITION Pyrococcus furiosus DSM 3638, 12452 bp DNA linear RCT 25-FEB-2002  
 ACCESSION AE010265 AE00950  
 VERSION AE010265\_1 GI:18693813  
 KEYWORDS SOURCE  
 ORGANISM Pyrococcus furiosus DSM 3638  
 PYrococcus furiosus DSM 3638  
 Archaea; Eurarchaeota; Thermococcales;  
 Thermococaceae; Pyrococcus.  
 REFERENCE 1 (bases 1 to 12452)  
 AUTHORS Maeder, D.L., Weiss, R.B., Dunn, D.M., Cherry, J.L., Gonzalez, J.M.,  
 DiRuggiero, J. and Robb, F.T.  
 TITLE Divergence of the hyperthermophilic archaea Pyrococcus furiosus and  
 P. horikoshii inferred from complete genomic sequences  
 JOURNAL Genetics 152 (4), 1299-1305 (1999)  
 MEDLINE 99359404  
 PUBMED 10430560  
 REFERENCE 2 (bases 1 to 12452)  
 AUTHORS Robb, F.T., Maeder, D.L., Brown, J.R., DiRuggiero, J., Stump, M.D.,  
 Yeh, R.K., Weiss, R.B. and Dunn, D.M.  
 TITLE Genomic sequence of hyperthermophile, Pyrococcus furiosus:  
 implications for physiology and enzymology  
 JOURNAL Meth. Enzymol. 330, 134-157 (2001)  
 MEDLINE 21079003  
 PURMED 11240495  
 REFERENCE 3 (bases 1 to 12452)  
 AUTHORS Weiss, R.B., Dunn, D.M., Robb, F.T. and Brown, J.R.  
 TITLE The complete sequence of the Pyrococcus furiosus genome  
 JOURNAL Unpublished  
 REFERENCE 4 (bases 1 to 12452)  
 AUTHORS Weiss, R.B.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-FEB-2002) Human Genetics, University of Utah, 20  
 South 2030 East, Salt Lake City, UT 84112, USA  
 FEATURES source  
 1. 1-12452  
 /organism="Pyrococcus furiosus DSM 3638"  
 /strain="DSM 3638"  
 /db\_xref="taxon:186497"  
 gene  
 132..1013  
 /gene="PR1657"  
 132..1013  
 /gene="PR1657"  
 /note="Function Code: 13.1 Translation: Amino Acyl tRNA  
 Synthetases; (hiss)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hisdidyl-tRNA synthetase"  
 /protein\_id="RAL81781.1"  
 /db\_xref="GT:18893014"  
 /translati... "MWAKKSEEEVIIKIAETANKLRSIQFLWGYKEVVFPIIERYSPPEI  
 RKGTRFAVNNDIIFLYLSPDVPSRILKQFDQNAKYFYTAEVLDVKGWQASIEFTGG  
 KGLGMYLEALSIATTSLSBEGIYDIDGSIWELSGEYWPKTKEALIKRNFEI  
 ENPIPKDKKKWMEFLNRGRSGIEKLDITVSDIEPFRVIDLGTVWRPLPYTDI  
 FEIISPOVKPIIGGGEYEVRKGKAIGFVFYLNLSSLYRPDKDRARKVISEIDPKKAY  
 STAREMRIGIPVPEVRL"  
 1010..1624  
 /gene="PR1658"  
 1010..1624  
 /gene="PR1658"  
 /note="Function Code: 1.7 Amino Acid Biosynthesis:  
 gene  
 CDS  
 /transl\_table=11  
 /product="n-(5'-phospho-d-ribosylformimino)-5-amino-1-  
 (5'-phosphoribosyl-1'-4-imidazolecarboxamide isomerase"  
 /protein\_id="RAL81785.1"  
 /db\_xref="GT:18893819"  
 /translation="MKFRRTYPAIDMDGKVVRVLKGKKDEVVGDPLFAEKFSEYV"



Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,  
 Rappooli, R. and Ventur, J.C.  
 TITLE  
 JOURNAL  
 Submitted (17-MAR-2000) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA  
 FEATURES  
 source  
 /organism="Neisseria meningitidis MC58"  
 /strain="MC58"  
 /db\_xref="taxon:122586"  
 /note="serogroup: B"  
 gene  
 1. .11932  
 /gene="NMB1573"  
 CDS  
 102. .1097  
 /gene="NMB1573"  
 /note="similar to GB: M34930 SP: P21302 GB: X64860 PID: 150241  
 PID: 4487 percent identity: 97.58; identified by sequence  
 similarity; putative"  
 /product="ornithine carbamoyltransferase, catabolic"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ornithine carbamoyltransferase, catabolic"  
 /protein\_id="AAF41926.1"  
 /db\_xref="GI:7226821"  
 /translation=MNLANKRHLKLDFTEPEITAYLDLAAELKAAKKAGREIQRSTMKG  
 KIALIFEKSTTRPCCFAVARODAGSYTLPASOSIGHESIKDPAVGRGMYD  
 LEYGGFGQWVEELAKYAGCVPNGLTNEFPTQMLADALMRERSKPLNQTAFAY  
 GDAKYNMGNSLLTGAKLGMVRIGAPSOPLWPSBGIAKPLTENAH  
 EAVRNVDFTHTDWVMSCEPKWQERIDLKQYRVPPELMASGNPOKFHCLPAP  
 Ccomplement(1321. .2334)  
 /gene="NMB1574"  
 complement(1321. .2334)  
 /note="similar to GB: AEG000512 percent identity: 77.25;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ketol-acid reductoisomerase"  
 /protein\_id="AAF41927.1"  
 /db\_xref="GI:7226822"  
 /translation=MVOYDKADLSLIKGTKVATIGYSQDQHAANLKDGGVVNT  
 GLRGSSWMKKAERAGHVVKTVRAEAKREADVMMILLPDETMPAVYHATVANKREGATL  
 AFAIGFNVIVNQTPRADUDVIAVAPKPGHVTWSEKRGGVPSLAVVONSGRK  
 DIALSYAANGGKGGVETTEPDKVAVVQVLCVGGVVELIKGFFTTTEASYA  
 PEMAYFECHEMKLIVDIFEGGTIANMVISINNAEGEYVIGVGPVEVASSSEAMRIA  
 LKKRQTCGEYAKMFIQEGNNYASMTARRRNADHQVERVGAQURAMPWTANKLVQ  
 DKN"  
 complement(2414. .2707)  
 /gene="NMB1575"  
 complement(2414. .2707)  
 /gene="NMB1575"  
 /note="conserved hypothetical protein; identified by  
 Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAF41928.1"  
 /db\_xref="GI:7226823"  
 /translation="MSNKVILVTKPEYTHELAOFKELVKASRAEEGNNSYDLHQ  
 ERGKPNPNEVFEVWKSQAAIDHNASHQFQAFQVSQDGKTELETLVNEAD"  
 complement(2774. .3265)  
 complement(2774. .3265)  
 /gene="NMB1576"  
 /note="similar to PID: 147816 GB: U00096 PID: 1786266 percent  
 identity: 79.75; identified by sequence similarity;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="acetoacetate synthase III, small subunit"  
 /protein\_id="AAF41929.1"  
 /db\_xref="GI:7226824"  
 /note="similar to PID: 147816 GB: U00096 PID: 1786266 percent  
 identity: 79.75; identified by sequence similarity;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAF41933.1"  
 /db\_xref="GI:7226828"  
 /translation="MRHILSVLIEENSGAMSRSVWGLFSARDYNIDSLSLAVPTEKTL"









VERSION	AE000963.1	GI:	2689286
KEYWORDS	Archaeoglobus fulgidus	SOURCE	
ORGANISM	Archaeoglobus fulgidus	REFERENCE	
/product="preprotein translocase SecA subunit"		Archaeoglobales: Archaeoglobaceae; Archaeoglobus	
/protein_id="CA8B4963_1"		Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;	
/db_xref="GI:7360377"			
/db_xref="SPREMBL:Q0JTK7"			
/translation="MLNTIAKRRIGGERNDRLLKQYRSVARINAFLFQMOALSADLQ			
AKTAEKFKQLADGOTDGLPPEAVCREASRRTLGMRHEDVOLGGMVLDGKIAEM			
KYEGKVLIVATVLAVALAGKGVHVVTNDYLASRDAGTIMEPLNFGILTGVISD			
MQPFDRONAYAADTIVYTGNNERGFYLDNNAVTDQDKVARELNFAVDEVDSTIDE			
ARTPLITSGOADDNTOLYOTINTVPHLVRGETEGEYDWWDERAQVILSEACHEH			
AQDLTQNLLEANSANALMHMALARHTFLKDHHVWIDCETVIVDE			
TGRLMSCRRWSGLHQVAEKGVEKRNLTFLASTTFTONFRFLYDVKLSMGTADE			
AFEFQSYTNLETFIPTPNRPYORKDFENDQIFRSSEEKEFEAVVKDIEBCHKRQGPVIL			
Query Match Score 8.2%; Score 87.6; DB 1; Length 326301;			
Best Local Similarity 47.4%; Pred. No. 2.5e-13;			
Matches 329; Conservative 0; Mismatches 359; Indels 6; Gaps 2;			
Qy 182 AAGGGTCTTGGATTTGGCTTAAGGAGAAAATTAGTAGTCTCGGTACCGTT 241			
Db 54377 AAGCTTACGTTCGGCTGACATCCGCACATCGCCGACATCGCCGGCACCGCT 54436			
Qy 242 CGGAGGAACTCATACTACTCTCAATAGCTATAAGGGAACCTATAACCGTTACA 301			
Db 54437 CGGAGGAACTGATAGTCTCGGTACCGCTAACCGGGCGGGGAAATGTTGG 54496			
Qy 302 TACCTGTCCCACCTTCCACATGCGGATAATGGCGAACGTTCTCGGAAGACCCCTG 361			
Db 54497 CAGGGAAACGAGCTTCGATGCCACACGGCGCTACGGCATCGATGATATG 54556			
Qy 362 TAAGGTAACTGACGAAACTTGATATAGCTAGAAAGAATGATTAATAG 421			
Db 54557 TCGGGTTCACTTACACGGAGATTTCACCTCACCTGCGCGCTCTCGAACCGTCA 54616			
Qy 422 AGAAGAAAACCGCTTCGGGACTTTGCTTACCCAACACCCACGGGAACCTCT 481			
Db 54617 GGAAAGCACCCTCCGGCTACGCTTATGCCCTACCCACACCCACGGCTATGGT 54676			
Qy 482 TTTCAGGGAAATGAGGAGATAAGAACACG - --GGTGTTCGTGTATAGAGC 538			
Db 54677 TCACCGTGTGGCAATCGAGCCGTCATGGAGCTACAGGGATCTGCGTCGCG 54736			
Qy 539 AGCCCTACTATCATCTCCGGGAAACCTTCTCGGAAGACGCCCTCAAAGGGAGATA 598			
Db 54737 AAGCTTACCGGGCAATTCAACGGCACAGCTTCGCCAGGGCGACATTCAC 54796			
Qy 599 CGGTAGTTGAGACACTTCACAAATCGTATGGCGAGTTAGGGTAGGAAATTAA 658			
Db 54797 TCATGGTCATGGTACCGTCAGCAAACTGGTTGGCGACTGGTATGGCTATGCC 54856			
Qy 659 TAGGGAGGGAAATGTCCTCAGAAATACAGGTGAGACTCCCTCACCTGACCT 718			
Db 54857 CAGGGCTGCCGAAGTCATGGGAACTGCAATGGTAAATTAATGAAAC 54916			
Qy 719 ACCCCTCTCAGGTGAGGCAAAAGTCTCCATCGAGGAAGGAATCTTAATGGAA 778			
Db 54917 AACCTGACCTGACACGCCAACCTCGCCCT-GCAACACTACGGCATCATCTCGCA 54973			
Qy 779 AGATACAGGGGTGAAAGAGGAGATGACGAGAAATGAGAAATAGAAG 838			
Db 54974 ACATGGACACGCTGAAAGAACGAAACGGATGTCGCCGAATTGGCAAAATGCC 55033			
Qy 839 GAGTTGGGTTTCGAGTAACTCTT 872			
Db 55034 GTCTGACACCTTCCAACTGCAAT 55067			
gene			
CDS			
COMMENT			
FEATURES			
source			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS	2 (bases 1 to 2204)		
Kleck, H.P., Clayton, R.A., Tomb, J.-F., White, O., Nelson, K.E.,			
Ketchum, K.A., Dodson, R.J., Gwynn, M., Hickey, E.K., Peterson, J.D.,			
Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyprides, N.C.,			
Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,			
Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,			
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,			
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,			
Utterback, T., Cottrell, M.D., Spriggs, T., Artiach, P., Kaine, B.P.,			
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,			
Garrison, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,			
Woese, C.R. and Venter, J.C.			
The complete genome sequence of the hyperthermophilic,			
sulphate-reducing archaeon Archaeoglobus fulgidus			
Nature 390 (6658), 364-370 (1997)			
2 (bases 1 to 2204)			
Kleck, H.P., Clayton, R.A., Tomb, J.-F., White, O., Nelson, K.E.,			
Ketchum, K.A., Dodson, R.J., Gwynn, M., Hickey, E.K., Peterson, J.D.,			
Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyprides, N.C.,			
Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,			
Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,			
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,			
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,			
Utterback, T., Cottrell, M.D., Spriggs, T., Artiach, P., Kaine, B.P.,			
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,			
Garrison, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,			
Woese, C.R. and Venter, J.C.			
Direct Submission			
Submitted (15-DEC-1997) The Institute for Genomic Research, 9712			
Medical Center Dr, Rockville, MD 20850, USA			
REMARK	In order to show the genes in ascending order on the genome, the		
origin of this version has been moved by TIGR to position 203570			
of the original version and the opposite strand is shown from the			
original version.			
On Dec 16, 1997, this sequence version replaced gi:2648504.			
On Dec 16, 1997, this sequence version replaced gi:2648504.			
COMMENT	Location/Qualifiers		
JOURNAL			
REMARK			
source			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
Kleck, H.P., Clayton, R.A., Tomb, J.-F., White, O., Nelson, K.E.,			
Ketchum, K.A., Dodson, R.J., Gwynn, M., Hickey, E.K., Peterson, J.D.,			
Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyprides, N.C.,			
Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,			
Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,			
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,			
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,			
Utterback, T., Cottrell, M.D., Spriggs, T., Artiach, P., Kaine, B.P.,			
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,			
Garrison, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,			
Woese, C.R. and Venter, J.C.			
Direct Submission			
Submitted (15-DEC-1997) The Institute for Genomic Research, 9712			
Medical Center Dr, Rockville, MD 20850, USA			
REMARK	In order to show the genes in ascending order on the genome, the		
origin of this version has been moved by TIGR to position 203570			
of the original version and the opposite strand is shown from the			
original version.			
On Dec 16, 1997, this sequence version replaced gi:2648504.			
On Dec 16, 1997, this sequence version replaced gi:2648504.			
COMMENT	Location/Qualifiers		
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
Kleck, H.P., Clayton, R.A., Tomb, J.-F., White, O., Nelson, K.E.,			
Ketchum, K.A., Dodson, R.J., Gwynn, M., Hickey, E.K., Peterson, J.D.,			
Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyprides, N.C.,			
Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,			
Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,			
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,			
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,			
Utterback, T., Cottrell, M.D., Spriggs, T., Artiach, P., Kaine, B.P.,			
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,			
Garrison, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,			
Woese, C.R. and Venter, J.C.			
Direct Submission			
Submitted (15-DEC-1997) The Institute for Genomic Research, 9712			
Medical Center Dr, Rockville, MD 20850, USA			
REMARK	In order to show the genes in ascending order on the genome, the		
origin of this version has been moved by TIGR to position 203570			
of the original version and the opposite strand is shown from the			
original version.			
On Dec 16, 1997, this sequence version replaced gi:2648504.			
On Dec 16, 1997, this sequence version replaced gi:2648504.			
COMMENT	Location/Qualifiers		
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
Kleck, H.P., Clayton, R.A., Tomb, J.-F., White, O., Nelson, K.E.,			
Ketchum, K.A., Dodson, R.J., Gwynn, M., Hickey, E.K., Peterson, J.D.,			
Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyprides, N.C.,			
Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,			
Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,			
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,			
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,			
Utterback, T., Cottrell, M.D., Spriggs, T., Artiach, P., Kaine, B.P.,			
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,			
Garrison, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,			
Woese, C.R. and Venter, J.C.			
Direct Submission			
Submitted (15-DEC-1997) The Institute for Genomic Research, 9712			
Medical Center Dr, Rockville, MD 20850, USA			
REMARK	In order to show the genes in ascending order on the genome, the		
origin of this version has been moved by TIGR to position 203570			
of the original version and the opposite strand is shown from the			
original version.			
On Dec 16, 1997, this sequence version replaced gi:2648504.			
On Dec 16, 1997, this sequence version replaced gi:2648504.			
COMMENT	Location/Qualifiers		
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
Kleck, H.P., Clayton, R.A., Tomb, J.-F., White, O., Nelson, K.E.,			
Ketchum, K.A., Dodson, R.J., Gwynn, M., Hickey, E.K., Peterson, J.D.,			
Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyprides, N.C.,			
Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,			
Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,			
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,			
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,			
Utterback, T., Cottrell, M.D., Spriggs, T., Artiach, P., Kaine, B.P.,			
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,			
Garrison, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,			
Woese, C.R. and Venter, J.C.			
Direct Submission			
Submitted (15-DEC-1997) The Institute for Genomic Research, 9712			
Medical Center Dr, Rockville, MD 20850, USA			
REMARK	In order to show the genes in ascending order on the genome, the		
origin of this version has been moved by TIGR to position 203570			
of the original version and the opposite strand is shown from the			
original version.			
On Dec 16, 1997, this sequence version replaced gi:2648504.			
On Dec 16, 1997, this sequence version replaced gi:2648504.			
COMMENT	Location/Qualifiers		
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
Kleck, H.P., Clayton, R.A., Tomb, J.-F., White, O., Nelson, K.E.,			
Ketchum, K.A., Dodson, R.J., Gwynn, M., Hickey, E.K., Peterson, J.D.,			
Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyprides, N.C.,			
Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,			
Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,			
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,			
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,			
Utterback, T., Cottrell, M.D., Spriggs, T., Artiach, P., Kaine, B.P.,			
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,			
Garrison, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,			
Woese, C.R. and Venter, J.C.			
Direct Submission			
Submitted (15-DEC-1997) The Institute for Genomic Research, 9712			
Medical Center Dr, Rockville, MD 20850, USA			
REMARK	In order to show the genes in ascending order on the genome, the		
origin of this version has been moved by TIGR to position 203570			
of the original version and the opposite strand is shown from the			
original version.			
On Dec 16, 1997, this sequence version replaced gi:2648504.			
On Dec 16, 1997, this sequence version replaced gi:2648504.			
COMMENT	Location/Qualifiers		
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
Kleck, H.P., Clayton, R.A., Tomb, J.-F., White, O., Nelson, K.E.,			
Ketchum, K.A., Dodson, R.J., Gwynn, M., Hickey, E.K., Peterson, J.D.,			
Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyprides, N.C.,			
Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,			
Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,			
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,			
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,			
Utterback, T., Cottrell, M.D., Spriggs, T., Artiach, P., Kaine, B.P.,			
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,			
Garrison, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,			
Woese, C.R. and Venter, J.C.			
Direct Submission			
Submitted (15-DEC-1997) The Institute for Genomic Research, 9712			
Medical Center Dr, Rockville, MD 20850, USA			
REMARK	In order to show the genes in ascending order on the genome, the		
origin of this version has been moved by TIGR to position 203570			
of the original version and the opposite strand is shown from the			
original version.			
On Dec 16, 1997, this sequence version replaced gi:2648504.			
On Dec 16, 1997, this sequence version replaced gi:2648504.			
COMMENT	Location/Qualifiers		
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
Kleck, H.P., Clayton, R.A., Tomb, J.-F., White, O., Nelson, K.E.,			
Ketchum, K.A., Dodson, R.J., Gwynn, M., Hickey, E.K., Peterson, J.D.,			
Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyprides, N.C.,			
Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,			
Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,			
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,			
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,			
Utterback, T., Cottrell, M.D., Spriggs, T., Artiach, P., Kaine, B.P.,			
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,			
Garrison, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,			
Woese, C.R. and Venter, J.C.			
Direct Submission			
Submitted (15-DEC-1997) The Institute for Genomic Research, 9712			
Medical Center Dr, Rockville, MD 20850, USA			
REMARK	In order to show the genes in ascending order on the genome, the		
origin of this version has been moved by TIGR to position 203570			
of the original version and the opposite strand is shown from the			
original version.			
On Dec 16, 1997, this sequence version replaced gi:2648504.			
On Dec 16, 1997, this sequence version replaced gi:2648504.			
COMMENT	Location/Qualifiers		
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
Kleck, H.P., Clayton, R.A			



	REFERENCE
	AUTHORS
	JOURNAL
Db	9834 ATTCCCAATGCCCTCCTACACCCTCTACTCCATCATGCAATGTGCGGAGGCCATCCATC 9893
Qy	361 GTAAAGGTCACTGGAGAAACCTTGTAGACTTAGAAGAGTTGATAATA 420
Db	9894 AGCTTCCCCTTGTAGGGATACAGGTGA ----- TCCCTGAGTCATC 9938
Qy	421 GAGAAAGAAAACCGTCTCGGTACTTGCTTACCCAAACACCCACGGAAACCTC 480
Db	9939 GCCCTGGAAAGGGAAAGCTCTGTTTCTGCTCTCCACACCCACGGAAACAGC 9998
Qy	481 TTTCAGGGGAAGATGGAGATAGAACAGGGTTCTGGTATAGAGAA 540
Db	9999 CTGAGCAGAAACTGTCAGAAGTGCTGAGTGCTGAGTGCTGAG 600
Qy	541 GCCTACTATCATTACTCCGGAGAACCTTCTGGAGACGCCCTAAAGGGAGATAGC 10058
Qy	658 ATAGGAAGGGGAATCGCTCAGAAATTACAGGTGAGTCCTCCATTACACCAACTG 600
Db	10179 GTCTGCAGTGGAGATTGCTGAGCAATTAGAGGTAGGCTAGGCTGCCCCTGGTATAAGC 10118
Qy	718 TACCCCTCTAGGATGGCAAA---ATCGGTAGGCGAGTTAGGGTAGGGATTTA 657
Db	10298 AAGGGACAGA - -TCGTCAGGAGGGAGGGAGGGCTATGAGAACTGAAGAATTAGG 10355
Qy	838 GGATGTGAGGTTCCTGGAGTAAGCTTAACCTCTGCTTTCAGAACGCCTACCCGCC 897
Db	10356 TGGCTTGAGTTACCGTCAGGCCACTCTGCTGTAANGGACAAGAGGGC 10415
Qy	898 CACGAGTTATCGAGCTACTGAAASGATGCTCTGCTCAGAACSTATCTACATG 957
Db	10416 GTGGTAG-----AGAGCTGGCCAGAAGGGTTATGTGAGGGATGCAAGTGTATG 10469
Qy	958 GAGGACTCCAATGCGCTCAGGTAACCGTAAAGGAAACAAAGTT 1017
Db	10470 GGCTTGAGGGTCTGTATTCGCAATTACAGTGGTAAGCGAGGATATGACAGGTG 10529
Qy	1018 CTGGAACACTGGAGGAGATTAATCCCTTCAAGCTCTC 1061
Db	10530 ATTGAGGCCCTAAGAGAAATTAGAAGAGGTCTCAGCCTTAT 10573
RESULT 11	
AL646072/C	
LOCUS AL646072	188050 bp DNA linear BCT 07-DEC-2001
DEFINITION Ralstonia solanacearum GMI1000 chromosome, complete sequence;	
ACCESSION Segment 16/19	
AL646072 AL646052	
VERSION AL646072.1 GI:17429824	
KEYWORDS SOURCE Ralstonia solanacearum.	
ORGANISM Ralstonia solanacearum.	
BACTERIA Ralstonia solanacearum; beta subdivision; Ralstonia group;	
REFERENCE (bases 1 to 188050)	
AUTHORS Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arialt, M., Billaud, A., Brottier, P., Camus, J.C., Cattonico, L., Chandler, M., Choisne, N., Claude, I., Renard, C.C., Cunin, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schleifer, T., Siguer, P., Thebault, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J., and Boucher, C.A.	
TITLE Genome sequence of the plant pathogen Ralstonia solanacearum	
JOURNAL Unpublished	
REFERENCE 2 (bases 1 to 188050)	
AUTHORS Boucher, C.A.	
TITLE Direct Submission	
QY SUBMITTED (05-DEC-2001) Genoscope and CNRS UMR-8030. 2 rue Gaston Crémieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS, BP27, 31226 Castanet-Tolosan Cedex, France, Fondation Jean Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, IMGM CNRS 118 Route de Narbonne, F-31062 Toulouse Cedex 4, Genoscope and INRA URGV, 2 rue Gaston Crémieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27, F31226 Castanet-Tolosan Cedex, Laboratoire de Génétique Cellulaire INRA, BP27, F31326 Castanet-Tolosan Cedex, Christian.Boucher@ceolouse.inra.fr, http://R.solanacearum.html.	
COMMENT 1. 188050	
FEATURES source	
source /organism="Ralstonia solanacearum"	
/strain="GMI1000"	
/db_xref="taxon:305"	
gene 73..909	
gene /note="Product confidence : probable	
gene name confidence : hypothetical	
predicted by Codon usage	
predicted by Homology	
/product="PROBABLE TRANSMEMBRANE PROTEIN"	
/protein_id="CAD16519_1"	
/job_xref="GI:17429825"	
/translation="MIALLISMASVCGNSFAKTLFPAAGAAGTYTRITIGAILLA	
LWRPWRLRHDRDAGTIALGYTLASMNLFLYSLTRDPIGIAIEFPGVPLVAL	
SRALDEVVIGAVAGLIGVAGLIGVAGLIGVAGLIGVAGLIGVAGLIGVAGLIGV	
PAGQATISLMVAGLAIPIPEVVOAGPALAALPSLIVAGGLGVLSVASSVYSLENVAL	
HISGRFTSVLISLEPAIGALAGAVVHLHEHSARQWVIAVITAASAGCATARSRAA	
ER" complement="MIALLISMASVCGNSFAKTLFPAAGAAGTYTRITIGAILLA	
/gene="RSC2803"	
/note="RS00302"	
complement:R36..2024	
/gene="RSC2803"	
/function="miscellaneous; hypothetical/partial homology"	
/note="Product confidence : hypothetical	
Gene name confidence : hypothetical	
predicted by Codon_usage	
predicted by FrameD	
/codon_start=1	
/evidence=not_experimental	
/transl_table=11	
/product="HYPOTHETICAL PROTEIN"	
/protein_id="CAD16510_1"	
/db_xref="GI:17429826"	
/translation="MALEIESTFAQYDQARARFLDARARGLATERAVHPHALGPAG	
EPUSIDPACFLDFDRPGWLLWTSINGVECGSGQVGVLJLDDALFARAGVALL	
LYHVNHYGFAHRYNEDNYDINRNSADFAVASANPAYLEVDPLLPWTPOAAN	
QALQQYLAMHEEALRDAMTGTGAYLPDMYGGATCWSTAQCACILSRRHADAP	
LAWIDLHNGLGAGHGRIFSCADPRELERATGTLAVIDRLLAAGSVSSVVEGLVD	
RAGLFPKEVKVILEFGTLEPMVMAOLRADHWLRHGTSPDQAATRKGLDAF	
YCDTPWKMVQAQARYAVLQLQARFSG"	
complement:R124..3326	
/gene="RrdB"	
CDS /EC_number="1.17.4.1"	
/function="small molecule metabolism; central intermediary	

```

<product>="PUTATIVE AMIDASE PROTEIN"
<protein_id>="CAD16_513_1"
<db_xref>="GI: 17429829"
<translation>"MPEAGVPRIGADGVCRRHPSNVNDADEPGLLIDILTVLN
SPLSPARGTDDFLPFANTLDGCAASIFPEQLRGVVSAFFIIRTGACYQFAPDARA
WHAGVDFEGTRCRNDTSIGEJECTDDQFTPAQYAAATVALVRACAYPIRAIAGH
SDTAPERTKTDGPGCFDWAIHLRLAKLEALFPYQGAAPA"
<gene>="pehr"
<complement>(742 . .9184)
<gene>="pehr"
<note>="RSC2807; RSC00298"
<complement>(742 . .9184)
</gene>
</gene>="pehr"
</product>
<function>="small molecule metabolism; global functions; global regulatory functions"
<note>="Product confidence : probable predicted by Codon_usage predicted by Homology predicted by FramedD"
<codon_start>1
<transl_table>11
<product>="PROBABLE RESPONSE REGULATOR TRANSCRIPTION
REGULATOR PROTEIN"
<protein_id>="CAD16_514_1"
<db_xref>="GI: 17429330"
<translation>"MSRQQFFRPMSKAAYVREPLVYDDEADLRELLEISLRRMGHD
VVLGGGLAAREALTRORRALVLDIMRLGDLGIDLVRLOSATADRIVATYAGSA
ENAEVALKAGAFDTIAPKSLDQSLRMLVLSRMLVLSRMLVLSRMLVLSRMLVLS
AAMQEVRSLLRLARSMAPVVISESGSSKERAARAIAHALSSSARPVAVNGCAIPE
NLMAEEFGVYKGAFTGASDQQFQANGIQLMLDEADLPLTMVQLLRAQERDV
VRKIGESRBDPVDRVCAQSIIQHNLVYAGPREDLRYRLVNLMPMTLRTREAED
PVLAQVLLBOLAAITGDSPLPKRULTRQALQCOLAYDPNVPNRELDNLJERAYAFAEGES
IDVPHLGAGLTGDIERSPLPKRULTRQALQCOLAYDPNVPNRELDNLJERAYAFAEGES
GPQDVGFGDDAAVTPVPMGLMPHPVTVVPEVEPAVIEPAPAMPAMSPVLDLPA

```

		ORIGIN	Query Match	Score	DB	Length
Db	167833	CCCTCCTCGGCCGCGCCAGGGCTTGTGGTCAGAGGGCTTGTGGTCAGAGGGCT 167834	6.8%	Score 72.2;	DB 1;	Length 1480;
Qy	545	ACTATCATTACTCCGGAACCTTCTGGAGAACGGGATACGGTAG 604	4.6%	Pred. No. 3 4e-09;		
Db	167833	ATAGCGCTTGCCAGCACGCTGGATGCCAACCTGTGCTGG 167774	0	Best Local Similarity	44.6%;	
Qy	605	TITTGAGGACACTTCAAAATGGTGGATTTAAGGTAGGGATTAAATAGGA 664	0	Mismatches	562;	Indels 12;
Db	167773	TAATGCGCAGTCCTCCAAGTGGCTGGGATCCCGCTAGGGCGGCC 167714	4	Matches	462;	Gaps 4;
Qy	665	ACGGGAAATTAACAGGTGAGCTTCAACCGTGACTTACCCCT 724	2	Conservative		
Db	167713	ACCGGAAGTGGTGTGGCGAGCTGCGAGTCAGCTGCTGA 167654	238	QY 2	TGATACCCAGGGATAAAGGAACCTTAAAGCTTACAAGACGGAGGTCACTTCGGCC 61	
Qy	725	CNTAGGTGATGCCAAAGTTCCTAACGGAGGAAGATCTTAATGAAAGATAC 784	239	Db 238	TGAGCTTCTCTTAACCAATTCCCTACAGACTTTCGGAGAGATAAAACAAAGGGCT 121	
Db	167653	CGAAGGCACCCCGCGTCTGGCT--GACAGCTGGCTGATGCCAGGCAG 167597	62	Qy 62	TGAGCTTCTCTTAACCAATTCCCTACAGACTTTCGGAGAGATAAAACAAAGGGCT 121	
Qy	785	AGGAGTTGTAAACAGCCGAGAAAGGGATGTAGCAGGAATAATGAAAGATGTT 844	298	Db 298	TGAGCTTCTCTTAACCAATTCCCTACAGACTTTCGGAGAGATAAAACAAAGGGCT 121	
Db	167586	CGACCCGTCGTCGCCAGGGTACCCGGTGTATGATCGCTGTCGCCAGCGGGCTGA 167537	357	Qy 298	TGAGCTTCTCTTAACCAATTCCCTACAGACTTTCGGAGAGATAAAACAAAGGGCT 121	
Qy	845	AGTTTTTCGAGTAAGGCTTAACCTCTTGTCTTCAAGAACGCCCTACGGG 904	357	Db 478	TGAGCTTCTCTTAACCAATTCCCTACAGACTTTCGGAGAGATAAAACAAAGGGCT 121	
Db	167536	CGTGTTCGCCAGGCCCAATTCTGCGCTGCGTCCCGATGCCG--CGGCC 167480	358	Qy 299	TGAGCTTCTCTTAACCAATTCCCTACAGACTTTCGGAGAGATAAAACAAAGGGCT 121	
Qy	905	TITATCAGGAGGTACTICAAGGGATGTCCTCGGATCTTACATGGAGGAC 964	538	Db 538	TGGCGGTGATCACCCCGAGTGTGCTGAGCTGGCTGCTGCGAGCTGGCTGCAT 597	
Db	167479	TGTTGCGACCGCTGCTCAAACCGGGCTCCCTGTCAAATAGCACCCTG 167420	359	Qy 359	TGTAAGGGTTCACTGGAGAAACTTGTATAGACTTAGAAGAAGTATGAAATTAA 418	
Qy	965	TCCA--AAGTGTCTAGGGTAAGCTAGGGAAACCCGAAGAAACAAAGTTCTGG 1.021	598	Db 598	TGCGGGCATTTCCCTGTGGACTCTCAGCTGACCTTCGGCATGCTGGGGCA 657	
Db	167419	TGCTGGCCAACACTCTCTGGCTGACGSTCAGGGAAACCCGAAGAAACGCGATTCTTG 1.67360	419	Qy 419	TGAGAAAGAAAACCCGTTCTCGGGTATTTGGTACCCAAACACCCACGGAAAC 478	
Qy	1022	AAGCACTGG 1030	658	Db 658	TGCGGGCATTTCCCTGTGGACTCTCAGCTGACCTTCGGCATGCTGGCACCC 717	
Db	167359	ATGCCCTTG 167351	479	Qy 479	TCTTTTCCAGGGAAAGATGGAGATAAAACACCCACGGAAAC -- AGGGGTGTTCTGTGTAATAG 535	
Db	7.18	TATCCCGGACCCATCTGAAAGCCATGTCGCGCATGTCGACCTTCGGCAT 777	777	Db 7.18	TATCCCGGACCCATCTGAAAGCCATGTCGCGCATGTCGACCTTCGGCAT 777	
Qy	536	ACGAAGCCTACTATCATTCATTCTCGGAGAAACCTTCTGGAAAGACGGCTCAAAGGGAG 595	596	Qy 596	ATACGCTAGTTTGGAGACTTCAAAATCGSTGTTACGCTGAGACTCCCTCAACGTGA 655	
Db	7.78	ATGAGCCATTACGCCCTAGTCAGGACCTTGTGCCATCACCTGGACGCC 837	837	Db 838	ACCTCCCTCTGCTGGCACCATGTCAGGACCTTGTGCCATCACCTGGACGCC 897	
Qy	597	TAAATGGAAAGGGAAATGTGCTAGAAATTAAAGGTGAGACTCCCTCAACGTGA 715	838	Db 838	CCTACCCCTCTGCTGGCACCATGTCAGGACCTTGTGCCATCACCTGGACGCC 775	
Db	8.98	TGGCGGGCCGGCTGGATTCTGGAACTGCTGGCTGGGGATGGCTGGGAGTC 897	897	Qy 716	CTTACCCCTCTGCTGGCACCATGTCAGGACCTTGTGCCATCACCTGGACGCC 775	
Qy	656	TAATGGAAAGGGAAATGTGCTAGAAATTAAAGGTGAGACTCCCTCAACGTGA 715	916	Db 916	CTTACCCCTCTGCTGGCACCATGTCAGGACCTTGTGCCATCACCTGGACGCC 775	
Db	8.99	TGGCGGGCCGGCTGGATTCTGGAACTGCTGGCTGGGGATGGCTGGGAGTC 897	916	Db 958	ATGACTCACCCAGCCAGTCGCTCTTACCT--GGCTCACCCGAGTGCTGAANG 1014	
Qy	776	AAAAGATCAGGAGGTGAAAGGATGTGACGGAGAAAGAAATAG 835	957	Db 1015	CCAGGGCGAAATTCTCGGTGTCAGGAGGACGGCTTACGCCAT--CCGTGCT 1071	
Db	10.02	GGGTCTTCTGGCTGGCCGAACTGCTGGCTGGGGATGGCTGGGGGG 1131	1131	Qy 836	AAGGAGTTGAGCTTTCGAGTAAGGTAACTGAAAGGATCTCCCTGCTGAGAATCTTACA 955	
Qy	896	CCCAGGGAGTTATCAGGAGTACTGAAAGGATCTCCCTGCTGAGAATCTTACA 955	1131	Db 1072	GGGTCTTCTGGCTGGCCGAACTGCTGGCTGGGGATGGCTGGGGGG 1131	
Db	11.32	CAGGGTGTCTGGCTAGGCCAGTCGCTTACAGGCCATTAAAGCCCTTACAGGCC 1191	1191	Qy 956	TGGAAGGAACTCCAAAATGGCTCAGGGTAACGGTAAAGGAAACACAAAGT 1015	
Qy	956	TGGAAGGAACTCCAAAATGGCTCAGGGTAACGGTAAAGGAAACACAAAGT 1015	1015	Db 11.92	ACCCCCCTCTGGCAATATCTACGGTGTGGCAACTCCGGTAAATGACCGCT 1251	
misc_feature		/db_xref="taxon:920"				
		/note="similar to histidinol-phosphate aminotransferase;"				
		EC-hisC"				
BASE COUNT	261 a	501 c	426 g	292 t		



gene	6005.	7018	/gene="hydgb-2"				
CDS	6005.	7018	/note="CT1249"				
	6005.	7018	/gene="hydgb-2"				
	/codon_start=1		/codon_start=1				
	/transl_table=11		/transl_table=11				
gene	product="hydrogenase/sulfur reductase, beta subunit"		/product="hydrogenase/sulfur reductase, beta subunit"				
	/protein_id="AAM72419_1"		/protein_id="AAM72419_1"				
	/db_xref="GI: 21647225"		/db_xref="GI: 21647225"				
	/translation="MTRILLRKNDLDECLAAWQAKGFSVSLAPVKRHEMSCFGEVOKSGD		/translation="MTRILLRKNDLDECLAAWQAKGFSVSLAPVKRHEMSCFGEVOKSGD				
	MALDVLMLERTIKDQFQPQPEPLIYKIGKQOIDSETMTTPEPEKKVKEFVRPCDASGL		MALDVLMLERTIKDQFQPQPEPLIYKIGKQOIDSETMTTPEPEKKVKEFVRPCDASGL				
	AIDDPFLFGWDKDDTWERREBKSVLVTIATDVTSVSKLPSDTSTGADVLMP		AIDDPFLFGWDKDDTWERREBKSVLVTIATDVTSVSKLPSDTSTGADVLMP				
	LSDGSWSOVAEYASDRQTLAVDTVSSVLLQAEAAQPAVPTAEEKVFEYKMEWALDN		LSDGSWSOVAEYASDRQTLAVDTVSSVLLQAEAAQPAVPTAEEKVFEYKMEWALDN				
	PESQFWKDKTIALRCVCGSCTFLCPICPCHCFD1QDECIDTYQG1RKRKNNDSCSFPLFMTM		PESQFWKDKTIALRCVCGSCTFLCPICPCHCFD1QDECIDTYQG1RKRKNNDSCSFPLFMTM				
	SGHNPRTNTQTWRWQRIMHKENYVRYGKFVNCSGCCGRCTRQCPVDMG1TETLQA1N		SGHNPRTNTQTWRWQRIMHKENYVRYGKFVNCSGCCGRCTRQCPVDMG1TETLQA1N				
	LPR"		LPR"				
gene	7045.	7881	/gene="hydg-2"				
CDS	7045.	7881	/note="CT1250"				
	7045.	7881	/gene="hydg-2"				
	/note="identified by match to pfam protein family HMM		/note="identified by match to pfam protein family HMM				
	PF00175"		PF00175"				
gene	/codon_start=1		/codon_start=1				
	/transl_table=11		/transl_table=11				
	/product="hydrogenase/sulfur reductase, gamma subunit"		/product="hydrogenase/sulfur reductase, gamma subunit"				
	/protein_id="AAM72480_1"		/protein_id="AAM72480_1"				
	/db_xref="GI: 21647226"		/db_xref="GI: 21647226"				
	/translation="MLYSPFPMRVYSKRAEAQGVNLKLREFVQDHEFFKANYRTGM		/translation="MLYSPFPMRVYSKRAEAQGVNLKLREFVQDHEFFKANYRTGM				
	FGLGYFGEGESTFCVASYPEIRKEYLECTEROSGRVTTSLANTDAGDIVTERGPYGNR		FGLGYFGEGESTFCVASYPEIRKEYLECTEROSGRVTTSLANTDAGDIVTERGPYGNR				
	FPFFEFSKKNLFIAGQIALPPTRSVLIWSCLDQREKYYRDVTFIVYGARTVADLVYKIEL		FPFFEFSKKNLFIAGQIALPPTRSVLIWSCLDQREKYYRDVTFIVYGARTVADLVYKIEL				
	DLWKRQDVRILVLTDPGEPTPDWQDHVGFTPYTLEQQAAPSPEINTAVLCGPPIIMTKF		DLWKRQDVRILVLTDPGEPTPDWQDHVGFTPYTLEQQAAPSPEINTAVLCGPPIIMTKF				
	TILTALEKLGFTAENVYTILENRMKCGTIGKCSRNCNVGSIYICKEGPVFTADEVQAMQA		TILTALEKLGFTAENVYTILENRMKCGTIGKCSRNCNVGSIYICKEGPVFTADEVQAMQA				
gene	7887.	8000	/gene="CT1251"				
CDS	7887.	8000	/note="CT1251"				
	7887.	8000	/note="identified by Glimmer2; putative"				
	/codon_start=1		/codon_start=1				
	/transl_table=11		/transl_table=11				
	/product="hypothetical protein"		/product="hypothetical protein"				
	/protein_id="AAM72481_1"		/protein_id="AAM72481_1"				
	/db_xref="GI: 21647227"		/db_xref="GI: 21647227"				
	/translation="MKNDDTKRGKRSFLCKSFLILHDHPHENNILLIRFRHH"		/translation="MKNDDTKRGKRSFLCKSFLILHDHPHENNILLIRFRHH"				
gene	8028.	8591	/gene="pth"				
CDS	8028.	8591	/note="CT1252"				
	8028.	8591	/note="CT1252"				
	/gene="pth"		/gene="pth"				
	/note="identified by match to pfam protein family HMM		/note="identified by match to pfam protein family HMM				
	PF01195"		PF01195"				
gene	/codon_start=1		/codon_start=1				
	/transl_table=11		/transl_table=11				
	/product="peptidyl-tRNA hydrolase"		/product="peptidyl-tRNA hydrolase"				
	/db_xref="GI: 21647228"		/db_xref="GI: 21647228"				
Query	Match	6.5%	Score 68.8;	DB 1;	Length 11265;		
Best Local Similarity		45.4%	Pred. No. 3.8e-08;				
Matches	447;	Conservative	0;	Mismatches 522;	Indels 15;	Gaps 5;	
y	y	61	GTCAAGGCTTCCTCTAACCAAATCCCTAACGACTTCCGGAGATAAACAAAGGCC	120			
b	b	11070	GTCAAGGCTTCACCAAACCAAACCCCTCGATCGCCCTCGTGCCTGAGCAAGATT	11011			
y	y	121	TTAGAAGAATTAAAAAGGTTCCCTTGACAAAATACCCGAGACCCGAAGCGAAGAGTTA	180			

Db	11010	CTGGACCACTGGTCCACCGTATGGGTTAAGGAAAGAAAATTAGTCGGTACCGGGGC	10953
Qy	181	AAAGCGGTCTTGGGATTTTTCGGGTTTAAGGAAAGAAAATTAGTCGGTAAACGGT	240
Db	10950	ATGGGGCTATGGATCTGGTTCGGCTGAACGGAGTTGGTATCATGCAATGGT	10891
Qy	241	TCGGACGAACTCATATACTACCTCTCATAGTATAGTGAACCTTACATACCGTTAAC	300
Db	10890	TCGACAGATGCTACACCATCTTCATGGCTGCTGGGCTCTGGCAGGGT	10831
Qy	301	ATACCTGTCACCTTCCCATTGAGATAAGTGGAAAGTCTGGAAAGACCCCTC	360
Db	10830	ATTCCAGACSCCTTCGTTTCGCTTAAGGACAAGCTTGACAGCTTCAGGGGT	10771
Qy	361	GTAAGGTTCAACTGGGAAACATTGATATAGCTTAAAGGATTTAGTATTAAATA	420
Db	10770	GTCGAGTGGCCGATGTCATGCGATCTCTCTTACGATGTCGATGCAAGGCC	10711
Qy	421	GAGAAAACACCGGTTCTGGTACCTTACCCAAACACCCACCGGAAACCTC	480
Db	10710	AGGCCGAGAGGTGATTATGGCTCTCCACGGGAAACATCAGGACGAGTCG	10651
Qy	481	TTTTCAGGGAAAGATGGAGGATAAGAACGGGTGTTCTGTG -- -TAATAGAC	537
Db	10650	CTGTCACACAGAGATGTCGAAACTGGCACGCTTGCGACGCTATGCTGCGAT	10591
Qy	538	GAAGCTACTATCATTTACTCCGGAGAACCTTCTGGAGACGGCTCAAAGGAAGAT	597
Db	10590	GAAGGGTAGTCGACTTTTCGGCTGCAATCGGCGCTGACCTGATCACCG	10531
Qy	598	A -- -CGTAGTTGGAGACCTCAA -- -ATCGGTATGGCGGTTAAGGTAAGG	651
Db	10530	AACCTGATCTGGTGGCCACCATGTCAGGGCTTGCCTGGGAATGCGTATCGSC	10471
Qy	652	ATTTTAATGGAAAGGGAAATCTCTCAGAAATTAACTGGTGAACCTCCCTCAAC	711
Db	10470	TTTGCATTGCGAAATCACTTCAGGGGAAATCTCCAGGCAAGATTCCTGGCG	10411
Qy	712	GTGACCTACCCCTCAGGGTATGGCCAAAGTCTCCTACGGGAAAGAAATTCTA	771
Db	10410	TGAGCCGACTGGGGAAATCACTTCATGGTGTGCT -- -GAAACTACTCACTGTG	10354
Qy	772	ATGGAAAAGATACAGGGGTTATGGCTGAGCAGGGAAAGGTACCGCAAATGAAAGAAA	831
Db	10353	ACCGAGCCGCTGGCASTACCTCTCCCGAGCGGGCATGAGGGAGCTGACAGAA	10294
Qy	832	ATAGAGGAGGTGGGTTTCGGAGTAAGCTAACTTCTGCTCATTTGCTGCG	891
Db	10293	ATTCCTGGATTCACCTTCAGGGGAAACCACTTCTGCTCATTTGCTGCG	10234
Qy	892	CCGGCCACCGAGGTTATCAGGGCTACTGGCTTCTGGCTCAAGAACGTATCT	951
Db	10233	GCATC -- -CGAGGTGTCAGGAAGCTCAAATGGGAGTGGCTTGGCAACGTGTCG	10177
Qy	952	TACATGGAAAGSACTCCAAACTGGCTTCAAGGTAAGGTAGGGAAACGGGAAAC	1011
Db	10176	GGATACCCGCTTATGGAGAACTGGCTTCAATGGTGGCTCAAGGGAGAACGAC	10117
Qy	1012	AAGTTCCTGGAAAGTGGACTGGAGG	1035
Db	10116	CGCCPTTGGAAACTGGTGAAGAANG	10093

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophytina; Magnoliophyta; eu dicots; core eu dicots;  
Rosidae; eurrosidae II; Brassicaceae; Arabidopsis; Arabidopsis  
1 (bases 1 to 1285)

**REFERENCE** Query Match Score 6.08; Score 64.2; DB 8; Length 1285;  
**AUTHORS** Best Local Similarity 48.3%; Pred. No. 6 2e-07;  
Matches 180; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

Yamada, K., Banh, J., Chan, M. M., Chang, C. H., Chang, E., Dale, J. M.,  
Deng, J. M., Goldsmith, A. D., Lee, J. M., Onodera, C. S., Quach, H. L.,  
Tang, C., Toriumi, M., Wu, H. C., Yamamura, Y., Yu, G., Bowser, L.,  
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,  
Jones, T., Kamiya, A., Karlin, Neumann, G., Kawai, J., Kim, C., Lam, B.,  
Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C. J., Sakurai, T.,  
Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,  
Davis, R. W., Ecker, J. R. and Theologis, A.

**TITLE** Unpublished

**JOURNAL** Arbiddopsis Open Reading Frame (ORF) Clones

**REFERENCE** 2 (bases 1 to 1285)

Yamada, K., Banh, J., Chan, M. M., Chang, C. H., Chang, E., Dale, J. M.,  
Deng, J. M., Goldsmith, A. D., Lee, J. M., Onodera, C. S., Quach, H. L.,  
Tang, C. C., Toriumi, M., Wu, H. C., Yamamura, Y., Yu, G., Bowser, L.,  
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,  
Jones, T., Kamiya, A., Karlin, Neumann, G., Kawai, J., Kim, C., Lam, B.,  
Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C. J., Sakurai, T.,  
Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,  
Davis, R. W., Ecker, J. R. and Theologis, A.

**JOURNAL** Direct Submission

**COMMENT** Submitted (03-JUN-2002) Plant Gene Expression Center, 800 Buchanan  
Street, Albany, CA 94710, USA

The RIKEN Genomic Sciences Center (GSC) members carried out the  
collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN  
Arabidopsis Full-Length cDNA'): Seki, M., Narusaka, M., Ishida, J.,  
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,  
Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members constructed and  
sequenced the pun1 (ORF) clones using the RAFL cDNAs: Yamada, K.,  
Banh, J., Chan, M. M., Chang, C. H., Chang, E., Dale, J. M., Deng, J. M.,  
Goldsmith, A. D., Lee, J. M., Onodera, C. S., Quach, H. L., Tang, C. C.,  
Toriumi, M., Wu, H. C., Yamamura, Y., Yu, G., Bowser, L., Chen, H.,  
Cheuk, R., Jones, T., Karlin, Neumann, G., Kawai, J., Kim, C., Lam, B., Lin, J.,  
Miranda, M., Nguyen, M., Palm, C. J., Shinn, P., Southwick, A.,  
Davis, R. W., Ecker, J. R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally  
to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP  
/PGEC) contributed equally to this work as PIs.

**FEATURES**

**SOURCE**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**TITLE**

**JOURNAL**

**REFERENCE**

**AUTHORS**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Yamada, K., Liu, S.X., Phan, P.K., Banh, J., Banno, F., Dale, J.M., Goldsmith, A.D., Jiang, P.X., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin, Neumann, G., Kim, C., Koese, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M., Palm, C.J., Shin, P., Southwick, A., Tracy, S.E., Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC)

Contributed equally to this work as PIs.

Location/Qualifiers

1..1570

/organism="Arabidopsis thaliana"

/db\_xref="Taxon:3702"

/chromosome="5"

/clone="RAFL08-15-j07 (R11135)"

/note="This clone is in a modified pBluescript vector (CFLC-1) as a BamHI/XbaI insert.

ecotype: Columbia"

1..1570

/gene="At5g10330"

1..133

/gene="At5g10330"

134..1387

/gene="At5g10330"

/codon\_start=1

/evidence=experimental

/product="putative histidinol-phosphate aminotransferase"

/protein\_id="PAK92767\_1"

/db\_xref="GI:1529219"

/translation="MGVINVQGSPSFHSSESNLRSKRALKKPFCSLRNRYCAQS

SAAVDESKNTMGDSFIRHLRQLAAYQPLPPVLSQRLDPLDQNSLSESYIVVGCGADELLDLMRCV

LPDPGEKILIDQPTSMYVEDAVAGVVKLVPMPDFSLNDVIAEVVLEPKCIPV

TSPNNPDGSTSIEDDKLMPVLUVWPKVYENLTVRLTES

KRAGLAGLRYVGAPPLSIEYLWRAKOPYNVSAGEYAAALALSNGRYLEVDYRDALV

REERLFGLIKEVPLNPFTYSNSNILECTVSGMDAKKLKEDLAKMGVMVRHYNSQL

RGYVYVSAHPEHTDVLMECLQFY"

3' UTR

1..388..1570

/gene="At5g10330"

/note="compared to genomic sequence"

/replace="9"

444 a 298 c 344 g 484 t

m1sc\_difference

1553

/gene="At5g10330"

/note="compared to genomic sequence"

/replace="9"

444 a 298 c 344 g 484 t

BASE COUNT

ORIGIN

Query Match

6 0%; Score 64.2; DB 8; Length 1570;

Best Local Similarity 48.3%; Pred. No. 6.4e-07;

Matches 180; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

190 CTGGCGATTTCATGATGAGATAAGGAGAAAATTAGTTCTCGGTAAACGGTTTCGCGACGAA 249

527 CTGGCTCAAAGACTCTGGTTGGAGTCGAATACATCTGAGCTGTGCTGACGAA 586

250 CTGCATATCATCTGAATACATCTAGGTGAACTTACATACCCCTTACATACCTGTT 309

587 CTATGATGTTGATCATGATGAGTCGGCTGTTAGTCCTGGGAGAAGATTATAGATGTCCT 646

310 CCACACCTTCCATGATGAGATAAGTGAAGCTCGGAAGCCCCTGTAAGGT 369

647 CCAACTTCTCATGATGATGTGTTGATGCGGTGTAATGAGCTTAAGTT 706

370 CAACTGGACGAAACATTGATAGTAACTGTTAGAGATAAGTGAATGAAAGAA 429

707 CCAAGAAACCCCATTCAGCTGAATGAGCCGATCCTGAACTTGAAACTAGAA 766

430 AAACCGTCTCGGTACTTTGCTACCCAAACCCACGGGAAACCTCTTCCAGG 489

Db 767 AACCCAAATGATATCCAACTCTCCACAAATCCAGGGAGTATCATAGCGAG 826

Qy 490 GGAAAGATTGAGGAGATAAGAAACACGGGTTTCTGTTAATGACGAGCTACTAT 549

Db 827 GACGATCTGTTGAGATTCTGAAATGCCAAATCTGTTCTGATGAGCTTACATC 886

Qy 550 CATTACPCGGAG 562

Db 887 GAATTCTCAGGGAG 899

Search completed: April 15, 2003, 15:15:13  
Job time : 3628 secs

1

1

ALIGNMENTS

11:	/SIDS2/gcdata/geneseq/geneseq -emb1/NA1990.DAT:*		RESULT 1
12:	/SIDS2/gcdata/geneseq/geneseq -emb1/NA1991.DAT:*	AAT/8778	
13:	/SIDS2/gcdata/geneseq/geneseq -emb1/NA1992.DAT:*	ID	AAT/8778 standard; DNA, 1065 BP.
14:	/SIDS2/gcdata/geneseq/geneseq -emb1/NA1993.DAT:*	XX	
15:	/SIDS2/gcdata/geneseq/geneseq -emb1/NA1994.DAT:*	AC	AAT/8778;
16:	/SIDS2/gcdata/geneseq/geneseq -emb1/NA1995.DAT:*	XX	
17:	/SIDS2/gcdata/geneseq/geneseq -emb1/NA1996.DAT:*	DT	18-FEB-1998 (first entry)
18:	/SIDS2/gcdata/geneseq/geneseq -emb1/NA1997.DAT:*	XX	
19:	/SIDS2/gcdata/geneseq/geneseq -emb1/NA1998.DAT:*	DE	Aquifex histidinol-phosphate aminotransferase VF5/HPA DNA.
20:	/SIDS2/gcdata/geneseq/geneseq -emb1/NA1999.DAT:*	XX	
21:	/SIDS2/gcdata/geneseq/geneseq -emb1/NA2000.DAT:*	KW	Histidinol-phosphate aminotransferase; VF5/HPA, chiral compound
22:	/SIDS2/gcdata/geneseq/geneseq -emb1/NA2001A.DAT:*	KW	
23:	/SIDS2/gcdata/geneseq/geneseq -emb1/NA2002.DAT:*	XX	OS Aquifex sp. strain VF5
24:			

No. 15 The number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SYMPTOMS

SOMALIA								97WO-US010		
Result	No.	Score	Query Match	Length	DB	ID	Description	PF	21-JAN-1997;	
1	1065	100.0	AAT78778	18	AAT78778		Aquifex histidinol	XX		
2	90.8	8.5	AAA91484	21	AAA91484		N. meningitidis pa	XX		
3	90.8	8.5	AAE26111	21	AAE26111		Neisseria meningitidis pa	PR	08-MAY-1996;	96US-064655
4	90.8	8.5	AAA91489	21	AAA91489		N. meningitidis pa	PR	09-FEB-1996;	96US-059911
5	60.2	5.7	AAV21209	19	AAV21209		Methanococcus jann	XX		
6	46.2	4.3	ABK73007	24	ABK73007		Bacillus licheniformis	PI		
7	42.2	4.0	AAT78780	18	AAT78780		Ammonifex histidin	DR		
8	37.6	3.5	ABA28180	22	ABA28180		Human foetal liver	DR		
9	37.6	3.5	ABA28180	22	ABA28180		Probe #6646 for ce	XX		

PT - used for producing enzymes, and hybridisation probes for a cDNA or  
PR genomic library  
XX  
PS Claim 1; Fig 7; 95PP; English.  
XX  
CC This genomic DNA encodes histidinol-phosphate aminotransferase  
CC VFS5/HPA (see AAW24254) of *Aquifex* VFS5, a strictly  
CC chemolithoautotrophic, marine eubacterium which grows optimally at  
CC 85°C and pH 6.8 in high salt medium. The VFS5/HPA sequence  
CC can be amplified from a pBluescript vector that contains the DNA by  
CC PCR (see AAT8794-95). The DNA can be used for recombinant  
CC production of the aminotransferase and to develop hybridisation  
CC probes. Claimed thermostable transaminases and aminotransferases  
CC (AAW24248-57) can be produced from native or recombinant host cells  
CC for use with L- and D-amino acids for production of optically  
CC pure chiral compounds used in the pharmaceutical, agricultural and  
CC other industries. A method is claimed for transferring an amino  
CC group from an alpha-keto acid to an alpha-keto acid using a claimed  
CC enzyme.  
XX Sequence 1065 BP; 344 A; 220 C; 252 G; 249 T; 0 other;  
Query Match 100.0%; Score 1065; DB 18; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1 ATGATACCCCAGAGGATTAAAGGAACTTGAACTGTAAAGACGGTCACTCCGGCTCC 60	Db	721 CCCTCTCAGGTATGGCAAAGGTTCTCCACGGAGAAGAACTTCATGGAAAG 780
Db	1 ATGATACCCAGAGGATTAAAGGAACTTGAACTGTAAAGACGGTCACTCCGGCTCC 60	QY	781 ATACAGGAGGTCTGAAACAGCGGAAAGGATGACGAACTGTAAAGAAATAGAGGA 840
QY	61 GTCAAGGCTTTCMCTTAAGGAATTCGCCCTACTGACTTTCGGAGGATAAACAAAGGCC 120	Db	781 ATACAGGAGGTCTGAAACAGCGGAAAGGATGACGAACTGTAAAGAAATAGAGGA 840
Db	61 GTCAAGGCTTTCMCTTAAGGAATTCGCCCTACTGACTTTCGGAGGATAAACAAAGGCC 120	QY	841 GTTAGGGTTTCGGACTAAGCTTTCGGAGGATAAACAAAGGCC 900
QY	121 TAGAGAAATTAAAAGGTTCCCTGAAACAATAACCCAGAACCTGGTAAAGGAAGAGTTA 180	Db	841 GTTAGGGTTTCGGACTAAGCTTTCGGAGGATAAACAAAGGCC 900
Db	121 TAGAGAAATTAAAAGGTTCCCTGAAACAATAACCCAGAACCTGGTAAAGGAAGAGTTA 180	QY	901 GAGTTTATCAGGGCTACTGAAAGGTTTCGGAGGCTTACAGCTCTTAA 1065
QY	181 AAAGGGTTCTGGGATTTTTCGGGATTTTTCGGCTTAAGGAAGAAATTAGTCTCGGTAAACGGT 240	Db	901 GAGACTCCAAAGGTGCTTCAGGTAAAGGTTAGGTAAGGAAACACAAAGTTCTG 960
Db	181 AAAGGGTTCTGGGATTTTTCGGGATTTTTCGGCTTAAGGAAGAAATTAGTCTCGGTAAACGGT 240	QY	961 GGACTCCAAAGGTGCTTCAGGTAAAGGTTAGGTAAGGAAACACAAAGTTCTG 1020
QY	241 TCGGACGAACTTACATACCTCTCAATAGCTTACATGGTAAAGGAAATTAGTCTCGGTAAACGGT 300	Db	961 GGACTCCAAAGGTGCTTCAGGTAAAGGTTAGGTAAGGAAACACAAAGTTCTG 1020
Db	241 TCGGACGAACTTACATACCTCTCAATAGCTTACATGGTAAAGGAAATTAGTCTCGGTAAACGGT 300	QY	1021 GAAGCACTGGAGAGTATAAAATCCCTTCAGCTCTTAA 1065
QY	301 ATACCTGTTCCACCTTCCCATSTACAGGATATACTGGCAAACTTCTCGGAACCCCTC 360	Db	1021 GAAGCACTGGAGAGTATAAAATCCCTTCAGCTCTTAA 1065
Db	301 ATACCTGTTCCACCTTCCCATSTACAGGATATACTGGCAAACTTCTCGGAACCCCTC 360	QY	1021 GAAGCACTGGAGAGTATAAAATCCCTTCAGCTCTTAA 1065
QY	361 GTAANGTTCAACTGGACCAAACTTGATAGCTTAACTGATTAAGAAAGAACTTGAATAATA 420	Db	1021 GAAGCACTGGAGAGTATAAAATCCCTTCAGCTCTTAA 1065
Db	361 GTAANGTTCAACTGGACCAAACTTGATAGCTTAACTGATTAAGAAAGAACTTGAATAATA 420	QY	1021 GAAGCACTGGAGAGTATAAAATCCCTTCAGCTCTTAA 1065
QY	421 GAGAAAGAAAACCGGTTCTGGTACTTGTACCCAAACACCCACGGGAAACCTC 480	Db	1021 GAAGCACTGGAGAGTATAAAATCCCTTCAGCTCTTAA 1065
Db	421 GAGAAAGAAAACCGGTTCTGGTACTTGTACCCAAACACCCACGGGAAACCTC 480	QY	481 TTTCAGGGAAAGATGAGGAGATAAGAACAGGGTTCTGGTATAGACGA 540
QY	541 GCCTACTATCAATTACTCGGAAAAACCTTCTGGTACCCAAACACCCACGGGAAACCTC 540	Db	481 TTTCAGGGAAAGATGAGGAGATAAGAACAGGGTTCTGGTATAGACGA 540
Db	541 GCCTACTATCAATTACTCGGAAAAACCTTCTGGTACCCAAACACCCACGGGAAACCTC 540	QY	541 GCCTACTATCAATTACTCGGAAAAACCTTCTGGTACCCAAACACCCACGGGAAACCTC 540
QY	601 STAGTTGGAGACACTTCAAAATCGGTTGCGACTTAAAGGTAGGGATTATA 660	Db	541 GCCTACTATCAATTACTCGGAAAAACCTTCTGGTACCCAAACACCCACGGGAAACCTC 540
Db	601 STAGTTGGAGACACTTCAAAATCGGTTGCGACTTAAAGGTAGGGATTATA 660	QY	661 GGGAAAGGGAAATCTCTAGAAATTAAAGGTGAGACTCCCTAACGTGACCTCA 720
QY	661 GGGAAAGGGAAATCTCTAGAAATTAAAGGTGAGACTCCCTAACGTGACCTCA 720	Db	661 GGGAAAGGGAAATCTCTAGAAATTAAAGGTGAGACTCCCTAACGTGACCTCA 720

CC sequences, which are all used in the exemplification of the present invention. The nucleic acid sequences, protein sequences, and antibodies against them, can be used in the manufacture of a composition. The composition can be used as a medicament (or in the manufacture of a medicament) for treating, preventing or diagnosing infection due to Neisserial bacteria. For example, some of the identified proteins could be components of vaccines against Meningococcus B; against all serotypes; and/or against all pathogenic Neisseriae. Identification of sequences from the bacterium will also facilitate production of biological probes, particularly organism-specific probes. Attempts to make efficacious Menigococcus B vaccines have failed mainly due to antigen tolerance. Multivalent vaccines have also been tried but none have successfully overcome antigenic variability. The provision of further, complete sequences may provide an opportunity to identify secreted or surface exposed proteins that may be presumed targets for the immune system and which are not antigenically variable or at least more conserved than other more variable regions.

XX Sequence 13732 BP: 3488 A; 3771 C; 3357 G; 3116 T; 0 other;

Query Match 8.5%; Score 90.8%; DB 21; Length 13732;  
Best Local Similarity 47.7%; Pred. No. 8 9e-17;  
Matches 331; Conservative 0; Mismatches 357; Indels 6; Gaps 2;

Qy 182 AAGCGGTTCCTGGGATTTTTCGGCTTAAGGAAGAAAATTAGTGTCTCGGTAACGGTT 241  
Db 12654 AAGCATTAGTCGTCGTCGACATTCGGACTTCGCACTCGGCAACGGCAACGGT 12713

Qy 242 CGGACGAACCTACATACCTCTCAATAGCTATAGGTGACTTACATACCGTTACA 301  
Db 12714 CGGACGAACTGATGATCACGTCATCACGATGCTGACCCGCAAACGGGGCGGAATGTGG 12773

Qy 302 TACCTGTTCACCTTCACATGTACAGATAAATGGCAAAAGTCTGGAAACCCCTCG 361  
Db 12774 CAGCGAACCCAGTTCTGCATATACCGCACAACGCCGCTGTAGGCATGGATTATG 12833

Qy 362 TAAGGTTAACCTGACGAAACCTTGTATAGCTAGAAAGAAGTATGGATTAAATAG 421  
Db 12834 TCGGGTTCACGAAACGGAGATTTCACCTCAACCTGCCGCGTCTCGAAGGCCCTCA 12893

Qy 422 AGARAGAAAAACCGTTACCCAAACACCCACGGAAACCTCT 481  
Db 12894 GGAAACACGCCCTGCCCAGCTTTCGCTTACCCCAACACCCACCGCTATGCT 12953

Qy 482 TTTCAGGGAAANGATTGA--GGAGATAAGAACAGGGTTTCTGTATAAGACG 538  
Db 12954 TCAGGGCGCGAAATCAGGGTCATCGAAACGGCTCATGCTGCGTGTGATGATG 13013

Qy 539 AAGCTTACTATCATTCATTACTCGCGGAAACCTTCTGGAAAGACCGCTCAAAGGAAGATA 598  
Db 13014 AAGCTTACGGCGATTCACCGCGACGCTTCCTGGCAAGCGAGGATTCACCAAC 13073

Qy 599 CGGTAGTTGAGGACACTTCAAAATCGGTATGGGAGTTAAGGTAGGGATTAA 658  
Db 13074 TGATAGTTAGCAACCTCAGCAARATGGTTTGCGGACTATCGTTATGGG 13133

Qy 659 TAGGAAAGGGAAATCGTCGAAATTAAAGGTGAGACTCCCTTAACGTTGACCT 718  
Db 13134 CAGCTGCCGAGTCATGGGAAACTGCAAAATTCGCGCTCAATATGAACC 13193

Qy 719 ACCCTCTCAGGTGATGGCAAAAGTCTCCACGGAGGAAGAATTCTTAATGAAA 778  
Db 13194 AATGGAGGCTGACCAACTCGCCCTGCGGCACTACGGCAATTATCTCTT--GCCA 13250

Qy 779 AGATACGGAGGTTAACAGAGCGGAAAGATGATGAGCAAAATAGAAAG 838  
Db 13251 ACATGACAGCTGAAAGCAACGGCAACTGCAATGGGAAATAATGCC 13310

Qy 839 GAGTTGAGTTTCGGAGTAAGGTAAACTACGGCAATTATCTCTT--GCCA 872  
Db 13311 GTCTGAACACCTTCAAGTCAGGAAACTCAT 13344

RESULT 3  
ID AAF21611 standard; DNA: 349980 BP.  
XX  
XX  
AC AAF21611;  
XX DT 13-MAR-2001 (first entry)  
XX DE *Neisseria meningitidis* B nucleotide sequence SEQ ID NO:112.  
XX KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; immunogenic; vaccine;  
XX diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
XX ds.  
XX OS *Neisseria meningitidis*.  
XX PN WO200066791-A1.  
XX PD 09-NOV-2000.  
XX PF 08-MAR-2000; 2000WO-US0505928.  
XX PR 30-APR-1999; 99US-0132068.  
PR 08-OCT-1999; 99WO-US23573.  
PR 28-FEB-2000; 2000GB-0004695.  
XX PA (CHIR ) CHIRON CORP.  
PA (GENO-) INST GENOMIC RES.  
XX PI Pizza M, Hickey E, Peterson J, Tettelin H, Ventter JC,  
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V,  
PI Frazer CM, Grandi G,  
XX DR WPI; 2000-647603/62.  
XX PT *Neisseria meningitidis* B full length genome sequence and open reading frames are used to detect, treat and prevent *Neisseria* infections -  
XX PS Claim 7; Appendix A; 692PP; English.  
XX  
The present invention describes the full length genome of *Neisseria meningitidis* B (NMB). The sequences in AAF21544 and AAF21607 CC represent fragments of the NMB genomic sequence, as the CC to AAF21613 represent fragments of the NMB genomic sequence, as the CC sequence was too long to go in a record on its own and of each CC sequences which overlap each other at the beginning and end of each CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at CC the beginning of AAF21607, the last 49980 bp of AAF21608, and so on). AAF21545 to AAF21588 encode the CC *Neisseria* proteins given in AAB5850 to AAB5859, and AAF21589 to CC AAF21606 represent PCR primers which are used in the exemplification of CC the present invention. The NMB genome and fragments from it have CC antibacterial activity, and can be used in vaccines and gene therapy. CC *Neisseria* nucleic acids, proteins and/or antibodies which binds to the CC proteins can be used in compositions for treating or preventing infection CC due to *Neisseria* bacteria or as a diagnostic reagent for detecting the CC presence of *Neisseria* bacteria or of antibodies raised to *Neisseria* CC bacteria. Computers, computer memory, computer storage medium or computer CC databases can be used in a search to identify open reading frames (ORFs) CC or coding sequences within the NMB genome. The DNA sequences provide CC further opportunities to find antigenic or immunogenic proteins which are CC more effective in vaccines than the outer membrane proteins currently CC used.  
XX Sequence 349980 BP; 87189 A; 93501 C; 84627 G; 84663 T; 0 other;  
Query Match 8.5%; Score 90.8; DB 21; Length 349980;  
Best Local Similarity 47.7%; Pred. No. 4-e-16;  
Matches 331; Conservative 0; N mismatches 357; Indels 6; Gaps 2;

QY	242	CGGACGAACTCATATACTACCTCTCAATAGCTTATAGTGAACTTTACATACCCGGTTTACA	301
Ddb	144679	CGGACGAACTGATACAGTTACAGGTCACTCAGTGTCAACGGGAAATGGGGAAATGTGG	144738
QY	302	TACCTGTTCCACCTTTCCTCATGTTACCTGTTACAGGATAAAGTGGCAAAGTTCCTGAAAGGCCCTCG	361
Ddb	144739	CAGCCGAACCCAGTTCTCATGTTACCTGAAAGCTGAAAGTGGAAATGGGAAATGTGGATTAGT	144798
QY	362	TAAGGTTCAACTGGACCAAACCTTGATATAGACTTAGAAAGAAGTATTGAAATTATAG	421
Ddb	144799	TGGGCTTCCACCTGAAAGGAGATTCAACCTGCCCCTCAACCCGCTTCAGAACGCCCTCA	144858
QY	422	AGAAAGAAAACCCTGTTCTGGGTGACTTTGCTTACCCAAACACCCCAAGGGAAACCTCT	481
Ddb	144889	GAAAACACGGCCCTGCCCCGTACCTTATGCCTTAACACCCCAACGGGTTAATGCT	144918
QY	482	TTTCCAGGGAAAGATTC - -GGAGATAAGAACAGGGGTGTTCTGTTAATAGACG	538
Ddb	144919	TCA CGGTGCGGAAATCAGCGTCACTGAACTTCTGGAAAGACGGCATCTCGTGTGATG	144978
QY	539	AGCCCTACTATCATTACUCCGGGAAACACTTCTGGAAAGACGGCTCAAAAGGGAAAGATA	598
Ddb	144979	AAGCCCTAGGGCATTCAACGGGACACGTTCTGCCAGGAGATCCCAACC	145038
QY	599	CGGTAGTTTGAGACACTTCAAAACGCTTCAAGGTTAGGGATTAA	658
Ddb	145039	TGATAGCTTCAACCTCAGGAAACGGTTTGGCTGAGCTGGGTATGGGG	145098
QY	659	TAGGGAAAGGGAAAATGCTCAGGAAATTAACAGGAGACTCCCTCAACTGACCT	718
Ddb	145099	CAGGCTGGCCCGAAGTCACTGGGAACCTCTCAGGGAAAGAATTCCTATGGAAA	145158
QY	719	ACCCCTCTAGGTGATGGCAAACACTTCTCTCAGGGAAAGAATTCCTATGGAAA	778
Ddb	145159	AATTGACCTGACCTGCAAACTGGCCACTACGGCAATTCTCT --GCCA	145215
QY	839	GAGTGAAGTTTCCGAGTAAGGCTTAATCTCTT	872

**RESULT 4**  
 AAA81489 standard; DNA; 837096 BP.  
 ID AAA81489  
 XX  
 AC AAA81489;  
 XX  
 AC AAA81489;  
 XX  
 DT 04-DEC-2000 (first entry)  
 XX  
 DE N. meningitidis partial DNA sequence gnm\_37 SEQ ID NO:37.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Menningococcus B; MenB; ds.  
 XX  
 Neisseria meningitidis.  
 OS  
 XX  
 PN WO200022430-A2.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 08-OCT-1999;  
 XX  
 PR 99WO-US23573.  
 XX  
 PR 09-OCT-1999;  
 PR 30-APR-1999;  
 XX  
 PA (CHIRON CORP.  
 XX  
 PA Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI

Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
Rappuoli R, Pizza M;

WPI; 2000-318079/27.

Isolated nucleotide sequences of *Neisseria meningitidis* which can be used in the diagnosis and treatment of *N. meningitidis* infection and other Neisseria infections, for example, *N.gonorrhoeae*.

Claim 7; Page 629-865; 1760pp; English.

The present invention describes methods of obtaining immunogenic proteins from *Neisseria* genomic sequences. *AAA81453* to *AAA82441* represent specifically claimed *Neisseria meningitidis* genomic DNA sequences; *AAA81360* to *AAA81303* and *AAB25620* to *AAB25663* represent *Neisseria* DNA sequences and their corresponding proteins; *AAA81234* to *AAA81259* and *AAA81304* to *AAA81321* represent PCR primers used in the isolation of *Neisseria meningitidis* DNA sequences; and *AAA81322* to *AAA81452* represent *Neisseria meningitidis* MenB polynucleotide ORE sequences, which are all used in the exemplification of the present invention. The nucleic acid sequences, protein sequences, and antibodies against them, can be used in the manufacture of a composition. The composition can be used as a medicament (or in the manufacture of a medicament) for treating, preventing or diagnosing infection due to *Neisseria* bacteria. For example, some of the identified proteins could be components of vaccines against *Meningococcus B*; against all serotypes and/or against all pathogenic *Neisseriae*. Identification of sequences from the bacterium will also facilitate production of biological probes, particularly organism-specific probes. Attempts to make efficacious *Meningococcus B* vaccines have failed mainly due to antigen tolerance. Multivalent vaccines have also been tried but none have successfully overcome antigenic variability. The provision of further, complete sequences may provide an opportunity to identify secreted or surface exposed proteins that may be presumed targets for the immune system and which are not antigenically variable or at least more conserved than other more variable regions.

Sequence 837096 BP; 207334 A; 227065 C; 197280 T; 2 other;

Db 209810 TGATAGTCCTACGCCACCCTAGCAAATCGGTTTGGGACTCGGTATGGTTATGCCG 209869  
 QY 659 TAGGGAGGGAAATCGTCAGAAAATTAAACAAGGTGACCTCCCTCAACGTGACCT 718  
 Db 209870 CAGGCTGCCCCGAAGTCATGGCAGACTTCAGTCAAATATGAAAC 209929  
 QY 719 ACCCCTCTCAGGTGATGCCAAAAGTCTCTCAGGGAGAAGAACCTCTATGGAA 778  
 Db 209930 AATTGAGCCTGACCACTCGCCCTGCAGCATTACCTCT - GACA 209986  
 QY 779 AGATACAGGAGGTTGTAACAGAGCAGAAAGGATGACGAAATAGAGG 838  
 Db 209987 ACATCGACCCGCTGAAAAGCAACGCCAGGGATTCGCCGAAATGAGC 210046  
 QY 839 GAGTGAGGTTTCGAGTAAGGCTAACTTCAT 872  
 Db 210047 GTCTGAACACCTTCAAGTCAGGCAAACTTCAT 210080

RESULT 5  
 ID AAV21209/c  
 ID AAV21209 standard: DNA; 1664376 BP.  
 XX AAV21209;  
 AC  
 XX DT 10-NOV-1998 (first entry)  
 XX DE Methanococcus jannaschii circular chromosome.  
 XX KW Methanococcus jannaschii; methanogenic archaeon; circular chromosome;  
 KW genome; autotrophic; extrachromosomal element; identification; ds.  
 OS Methanococcus jannaschii.  
 XX PN WO9307830-A2.  
 XX PD 26-FEB-1998.  
 XX PF 22-AUG-1997; 97WO-US14900.  
 XX PR 2-2-AUG-1996; 96US-0024428.  
 XX PA (GENO-) INST GENOMIC RES.  
 (UNI.) UNIV ILLINOIS FOUND.  
 (UTJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
 XX Bult CJ, Smith HO, Venter JC, White OR, Woese CR;  
 DR WPI; 1998-169145/15.  
 XX Complete genome sequence of methano-genic archaeon, Methanococcus  
 jannaschii - useful in identification of M. jannaschii genome  
 PT fragment  
 PS Claim 13; Page 152-585; 614pp; English.  
 XX  
 CC The present sequence represents the complete 1,66-megabase pair genome  
 CC sequence of the Methanococcus jannaschii circular chromosome. The  
 CC present invention describes M. jannaschii open reading frames from the  
 CC genome sequence. The invention also describes a computer based system  
 CC for identifying fragments of the M. jannaschii genome that are  
 CC homologous to target nucleotide sequences, comprising: (a) data storage  
 CC means comprising the nucleotide sequence of the 1664976, 58407 or 16550  
 CC bp sequence (see AAV21209, AAV2120 and AAV211), or a nucleotide  
 CC sequence at least 99.9% identical to it; (b) search means for comparing a  
 CC target sequence to the nucleotide sequence of the data storage means to  
 CC identify a homologous sequence; and (c) retrieval means for obtaining  
 CC the homologous sequence. The method, which is based on whole genome  
 CC random sequencing of an autotrophic archaeon M. jannaschii, the genome  
 CC of which consists of 3 physically distinct elements, a large circular  
 CC chromosome (the 1664976 bp sequence given in AAV21209), a large circular  
 CC extra-chromosomal element (the 58407 bp sequence given in AAV2120), and  
 CC a small circular extra-chromosomal element (the 16550 bp sequence given

CC in AAV21211), can be used in the identification of M. jannaschii genome  
 CC fragment.  
 XX SQ Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T; 101 other:  
 XX SQ Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T; 101 other:  
 Query Match 5.7%; Score 60.2; DB 19; Length 1664976;  
 Best Local Similarity 47.7%; Pred. No. 2.9e-06;  
 Matches 479; Conservative 0; Mismatches 493; Indels 33; Gaps 9;  
 QY 62 TCAAGGTTCCTCTAAGSAAATTCCCTACTAGACTTTCCGAGGATAAAACAAAGGGCCT 121  
 Db 586070 TAAACCTGGTCTCAATGAAAATCTGGGACCACCTCCAAAATAAAAGAAAAAATT 586011  
 QY 122 TAGANGAAATTAAAAGAGTTCCCTTGACAAATAACCCAGACCCGAGCAGAAAGCTAA 181  
 Db 586010 TAGATGAAATTGACAAATTCACCCAGACCCGAGCAGAAAGCTAA 585957  
 QY 182 AACCGGTTCTGGGTTTAAGGAAATAATTGAGGAAACATAATTGTTGAGGAGATG 585897  
 Db 585956 TGAAGAGTTAACAAATTGTAAGCTTGTAGGAAACATAATTGTTGAGGAGATG 585897  
 QY 239 GTTCGGAGAAGCTATACTACCTCTCAATAGCTTATAGGTGAACCTTACACCGTTT 298  
 Db 585896 GGCTGATGAGATTAGACAAATTTAACGTTTGTGTGTATGAGTGGATGAGGTTA 585837  
 QY 299 ACATACGTGTCCTCCACTCTTCCTCATGAGATAAGTGGCAAAGTCTGGAAACCCC 358  
 Db 585836 TAATTCCTAACATTCCAAATTTACCAATATAGAGTTCTAGCAACATCCAAATGCTAAA 585777  
 QY 359 TCGTAAGGTCAACTGGACAAAACCTTGTATAGACTTGTAGATGAAAGTGTGTTAA 418  
 Db 585776 TAAATATGCTAAATATGATAAGGAAAACACTTAAATGAAATGTGAAGTGTGCTTAA 585717  
 QY 419 TAGA--GAAGAAAMACCCCTTCTGGTACTTGTCTACCCAAACAGGGAA 475  
 Db 585716 ATAATATAACAGATAAAACGAAGGTATTTCCTGACTCCAAATACTCCACAGGAA 585657  
 QY 476 ACCTCTTTCCGGGGAAAGATTGA --GGGATAAGAAACAGGGTGTCTCTGTATA 532  
 Db 585656 ATATATAGAAAATAGAGAAGTAGAGGGTTATCATGAAACAGCCTTAACTGTGTTA 585597  
 QY 533 TAGACCAAGGCCCTACATCATTCACPCGGGAAACCTT-----CTGGAAAGACGCCCTCA 586  
 Db 585596 TTGACCATGCTACATGTGACTTAAAGGATAATGATGGACTCTAAAGCTTGTGTTA 585537  
 QY 587 AAAGGAAAGATACGTAAGTTGAGCACCTTCAGGAAATTCAGGAAATTC --GGATGGGAGSTAA 643  
 Db 585536 AAATATCATAATGTTATGTTAAAGAACCTTAAAGGCTTGTGTTA 585477  
 QY 644 GGTTAGGGATTTAATGGAAAGGGAAATTAACAGGTGACTCAGAAATTAAACAGGTGACCTC 703  
 Db 585476 GGTTACCGTACCTTGTGGTGTGGAAATTAAGGAGTTAAGCTTACGGGAGAG 585417  
 QY 704 CCTTCACGTGACCTTGTGGTGTGGAAATTAAGGAGTTAAGCTTACGGGAGAG 763  
 Db 585416 TATTTAGCTTAACAGGTTAAGTCAGTTAACGTTAACTGCTTAACTGATTAAAGAT - -AGAG 585360  
 QY 764 AATTCTTAATGGAAAGATAACGGAGGTGTAAGAGGAGATAAGGTGAGCTCAGCAGAA 823  
 Db 585359 AATTCTTGAAGATGTGTTAGAGGAAATTAAAGTAGAGATGCTCTAACATGGAT 585300  
 QY 824 TGAAGAAAATAGAGGAGTGGTTTCGCACTAAGCTAAGCTAAGCTAAGCTAAGCTAAGCTA 880  
 Db 585299 TGAAGAGTTAACGATAATAAAGTTATCTCTTCAGAACCTTATCTATGGTTGAT 585240  
 QY 881 GAACCCCTTACCCGCCAGAGGTTATCAGGAGCTACTGAAAGGGATGTCCTCGCTCA 940  
 Db 585239 TAAAAACAAATGAAAGCAAAAGAAATTTGTGAGGAACTTAAAGAGGTGTATGTTA 585180  
 QY 941 GGAAGGTTACCTTATGGAGG --ACTCACAAAGTGCCTCAGGGTAAAGCTTGTGTTA 585120  
 Db 585179 GAGATTGCACATCCTGTGTTAGGGATATAATGTTAGAGTCAATAGGGACGT 585120

Qy	998	CGGAAGAAAACAAGTTCTGGAGCACTTGAGGAGATAAA	1042	Qy	154	TACCCAGACCCSAAGGAAAGTAAAGGGTTCTGGGATTTTCGGCGTTAAG	213
Db	585119	TGAAAGGGTGAAGGTTTAAGATTTGAGGATTTAA	585075	Db	181	TATCCTGAGGGTATAGTGGSCTTGAAGAACAGTTGCCTCCATCCTGGTAAAT	240
<b>RESULT 6</b>							
ID	ABK73007	standard; DNA; 1085 BP.		Qy	214	GAAGAAATTAGTACTGGTAACTGGAACTCATACATACCTCTCAATAGCT	273
XX	AC			Db	241	GAAGGARATCATCTGGAAACGGAGCTACATCAAAATTCCCCTCT	300
DT	13-AUG-2002	(first entry)		Qy	274	ATAGGTGAACTTACATACCTGTTCCACCTTCCATGAGCAGATA	333
XX	Bacillus licheniformis	genomic sequence tag (GST) #298.		Db	301	TGTTGGACCTCCCTCAAACACTGTAAATGGGAATCGACATTCACAAAGCAC	360
DE	Differential gene expression; genomic sequenced tag; GST;			Qy	334	AGTCGAAAGTTCTGGAAAGACCCCTCGTAACTGACGAAACTTGTATA	393
KW	altered culture condition; environmental stress;			Db	361	AACCTGTATTCAGGAGCGGGAGTAGCTGAGTTCGCTGCATAACCGCTGCCAC	420
KW	physiological provocation; ds.			Qy	394	GACTGAAAGTATGAAATTAATAGAAAGAAAAACCGGTTCTCGGG	444
XX	Bacillus licheniformis.			Db	421	GACCTGTATGCAATGCTGAGGGATCGTGAACAGACGAAAGTTGCTGG	471
OS							
PN	WO200229113-A2.						
XX							
PD	11-APR-2002.						
XX							
PF	05-OCT-2001; 2001WO-US31437.						
XX							
PR	06-OCT-2000; 2000US-0880598.						
XX							
PR	27-MAR-2001; 2001US-279526P.						
XX							
PA	(NOVO ) NOVOZYMES BIOTECH INC.						
PA	(NOVO ) NOVOZYMES AS.						
XX							
PI	Berka R, Clausen IG;						
XX							
DR	2002-416584/44.						
XX							
PT	Monitoring differential expression of several genes in first Bacillus						
PT	cell relative to expression of same genes in one or more second						
PT	Bacillus cells, by using substrate containing Bacillus genomic						
PT	sequenced tag array -						
XX							
PS	Claim 4; SEQ ID NO 29B; 200pp; English.						
XX							
CC	The invention describes a method of monitoring differential expression of						
CC	genes in a first Bacillus cell relative to expression of the genes in						
CC	other Bacillus cells, comprising hybridising labelled nucleic acid probes						
CC	isolated from Bacillus cells to a substrate containing array of Bacillus						
CC	genomic sequenced tags (GST), examining the array, and determining						
CC	relative gene expression by an observed hybridisation reporter signal of						
CC	a spot in the array. The method is useful for measuring the expression of						
CC	genes in a first Bacillus cell relative to expression of the same genes						
CC	in one or more second Bacillus cells. The method is useful for monitoring						
CC	global expression of several genes from a Bacillus cell, discovering new						
CC	genes, identifying possible functions of unknown open reading frames and						
CC	monitoring gene copy number variation and stability. Monitoring changes						
CC	in expression of genes may be used to provide a representation of the way						
CC	in which Bacillus cells adapt to changes in culture conditions,						
CC	environmental stress or other physiological provocation. Extensive						
CC	follow-up characterisation is unnecessary, when one spot on an array						
CC	equals one gene or one open reading frame, since sequence information is						
CC	available. This sequence represents a genomic sequence tag (GST) used in						
CC	the method of the invention.						
CC	Note: The sequence data for this patent did not form part of the printed						
CC	specification, but was obtained in electronic format directly from WIPO						
CC	ftp.wipo.int/pub/published_pct_sequences.						
XX	Sequence 1085 BP; 302 A; 264 C; 272 G; 247 T; 0 other;						
Query Match	4.38;	Score 46.2;	DB 24;	Length 1085;			
Best Local Similarity	47.48;	Pred. No. 0.0014;					
Matches	138;	Conservative 0;	Mismatches 153;	Indels 0;	Gaps 0;		

XX Sequence 1092 BP; 217 A; 312 C; 357 G; 205 T; 1 other;  
 SQ Query Match Score 4.0%; Best Local Similarity 48.2%; Matches 119;保守性 0; Mismatches 128; Indels 0; Gaps 0;  
 Qy 138 GGTTCCCTTGAAACAAATACCGACCCCCGAAAGGGTAAAAGAGT 197  
 Db 168 GGACCACTGGCACCTTACCCAGGGTCAAAGTAGCTAGGTGGTAA 227  
 Qy 198 TTTTTTCGGGTAAAGGAGAAAATTATAGTTCTCGGTACGGGAACTCATATA 257  
 Db 228 GAAACTGGGATAGACCCGACAGCATATCCTGGGTCAGGGAACTCCA 287  
 Qy 258 CTACCTCTAAATAGCTATAAGGTGAACTTACATACCGTTAACATACCTGTTCCACCT 317  
 Db 288 GATGTCCTTGGCCCTGCTGCCGCCCCGACAGGTGGTCACTCTGGCTPACCT 347  
 Qy 318 TCCCCATGTAGCAGATAAGTGGCRAAGGTTCTCGGAACCCCTCGTAAGGTTCAACTGGA 377  
 Db 348 TCCCCGCTATGAGCCCCGGCACGGCTCATGGGGCTATCCGTAAGAATTTCAGGAAGTCCAGTA 407  
 Qy 378 CGAAAC 384  
 Db 408 GGACTAC 414

RESULT 8  
 ABA59603  
 ID ABA59603 standard; DNA; 588 BP.  
 XX AC ABA59603;  
 XX DT 01-FEB-2002 (first entry)  
 XX DE Human foetal liver single exon nucleic acid probe #7908.  
 XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
 XX OS Homo sapiens.  
 XX PN WO200157277-A2.  
 XX PD 09-AUG-2001.  
 XX PF 30-JAN-2001; 2001WO-US00669.  
 XX PR 04-FEB-2000; 2000US-0180312.  
 XX PR 26-MAY-2000; 2000US-0207456.  
 XX PR 30-JUN-2000; 2000US-0608408.  
 XX PR 03-AUG-2000; 2000US-0632366.  
 XX PR 21-SEP-2000; 2000US-0234987.  
 XX PR 27-SEP-2000; 2000US-0236339.  
 XX PR 04-OCT-2000; 2000US-0024263.  
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX PI Penn SC, Hanelz DK, Chen W, Rank DR;  
 XX DR WPI; 2001-483447/52.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human fetal liver -  
 XX PS Claim 1: SEQ ID NO 7908; 639pp + sequence listing; English.  
 XX The invention relates to a single exon nucleic acid probe for measuring human gene expression in a sample derived from human foetal liver. The single exon nucleic acid probes may be used for predicting, measuring and displaying gene expression in samples derived from human fetal liver. The present sequence is a single exon nucleic acid probe of the invention.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp://ftpt.wipo.int/pub/published-pct\\_sequences.i](http://ftpt.wipo.int/pub/published-pct_sequences.i).  
 CC Sequence 588 BP; 274 A; 71 C; 111 G; 132 T; 0 other;  
 XX SQ Query Match Score 3.5%; Best Local Similarity 50.6%; Pred. No. 0; Gaps 0;  
 XX Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
 Qy 680 CAGAAATTAAACAGGTGAGACTCCCTTCAACGTGACCTACCCCTCTCAGSTGATGGCAA 739  
 Db 309 CAGGCCCTAGCAGCTATGCTGGGACAACTAAATGTCACAATGTCAACTATGTTGTGGTGTGTTA 368  
 Qy 740 AAGTTCTCTACGGAGGAAAGAATTCCATGGAAANGATAACAGGAGGTGTAACAG 799  
 Db 369 CAGAGAGAGATGTTGGGGAGAAGGATTTGGGAAACAAAATAATGGCTGAAACT 428  
 Qy 800 AGGGAGAAAGGATGTAAGCACGAAATAGAAGGAGTTAGGTTTTCGGAGTA 859  
 Db 429 TCCCAAACTGTGAAAGACATAATAACATCCAGAAATTTCAGGAAGTCCAGTA 488

RESULT 9  
 ABA28180  
 ID ABA28180 standard; DNA; 588 BP.  
 XX AC ABA28180;  
 XX DT 23-JAN-2002 (first entry)  
 XX DE Probe #6646 for gene expression analysis in human heart cell sample.  
 XX KW Human; gene expression; heart; microarray; vascular system; probe; cardiovascular disease; hypertension; cardiac arrhythmia; congenital heart disease; ss.  
 XX Homo sapiens.  
 XX OS Homo sapiens.  
 XX PN WO200157274-A2.  
 XX PD 09-AUG-2001.  
 XX PF 30-JAN-2001; 2001WO-US00666.  
 XX PR 04-FEB-2000; 2000US-0180312.  
 XX PR 26-MAY-2000; 2000US-0207456.  
 XX PR 30-JUN-2000; 2000US-0608408.  
 XX PR 03-AUG-2000; 2000US-0632366.  
 XX PR 21-SEP-2000; 2000US-0234987.  
 XX PR 27-SEP-2000; 2000US-0236339.  
 XX PR 04-OCT-2000; 2000US-0024263.  
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX PI Penn SG, Hanelz DK, Chen W, Rank DR;  
 XX DR WPI; 2001-48889/53.  
 XX PT Single exon nucleic acid probes for analyzing gene expression in human hearts -  
 XX PS Claim 1: SEQ ID NO 6646; 530pp; English.  
 CC The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart. The present sequence is one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

XX Sequence 588 BP; 274 A; 71 C; 111 G; 132 T; 0 other;

SQ Query Match 3.5%; Score 37.6; DB 22; Length 588;

Best Local Similarity 50.6%; Pred. No. 0.45;

Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Query Match 3.5%; Score 37.6; DB 22; Length 588;

Best Local Similarity 50.6%; Pred. No. 0.45;

Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 680 CAGAAATTAAACAGGTGAGACTCCCTTCACCGTGACCTACCCCTCTCAGGTGATGGCAA 739

Db 309 CAGGCCATTAGCAATCTGGGACAACTTATGTTGTGTAGTTA 368

QY 680 CAGAAATTAAACAGGTGAGACTCCCTTCACCGTGACCTACCCCTCTCAGGTGATGGCAA 739

Db 369 CAGAAAGAGATGGTGGGGAGAACATAATGCTGAAACTTATGTCGAACTTATGTTGTAGTTA 368

QY 740 AAGTCTCTCCCTCACCGAGGAATTCCTATGGAAAGATACAGGAGTTGTAACAG 799

Db 369 CAGAAAGAGATGGTGGGGAGAACATAATGCTGAACTTATGTCGAACTTATGTTGTAGTTA 368

QY 800 AGCGAAAGAGATGGTACGCAAATGAGGAGTTGAGGTTCGGAGTA 859

Db 429 TCCCCAAATCTGATGAAAGACATAATAAACATCCAAAGAAATTAAAGAAATTAAGGAGTCCAAAGTA 488

RESULT 10  
AAK07876

ID AAK07876 standard; DNA: 588 BP.

XX AC AAK07876;

XX DT 05-NOV-2001 (first entry)

XX DE Human brain expressed single exon probe SEQ ID NO: 7867.

XX DE Human; brain expressed exon; gene expression analysis; probe;

XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

XX KW epilepsy; cancer; ss.

XX OS Homo sapiens.

XX PN WO2001517275-A2.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608048.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236159.

XX PR 04-OCT-2000; 2000GB-0024663.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR; 2001-488900/53.

XX PR Human genome-derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human bone marrow -

XX XX Example 4: SEQ ID NO: 8300; 658pp + Sequence Listing; English.

XX PS The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human

CC brain. They can be used to measure gene expression in bone marrow

CC samples which may enable the improved diagnosis and treatment of cancers

CC such as lymphoma, leukaemia and myeloma. The present sequence is one of

CC the probes of the invention.

XX SQ Sequence 588 BP; 274 A; 71 C; 111 G; 132 T; 0 other;

XX Query Match 3.5%; Score 37.6; DB 22; Length 588;

CC Best Local Similarity 50.6%; Pred. No. 0.45;

CC Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

XX QY 680 CAGAAATTAAACAGGTGAGACTCCCTTCACCGTGACCTACCCCTCTCAGGTGATGGCAA 739

XX Sequence 588 BP; 274 A; 71 C; 111 G; 132 T; 0 other;



**RESULT 14**

CC using human lung derived mRNA and for the study of lung diseases  
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberculous sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocytic  
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 CC and hyaline membrane disease. The present sequence is a single exon  
 CC probe of the invention.

CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC [ftp://wipo.int/pub/published\\_pct\\_sequences](ftp://wipo.int/pub/published_pct_sequences).

XX Human genome-derived single exon probe from lung SEQ ID No 8607.  
 XX Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; pulmonary histiocytosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; lymphangioleiomyomatosis;  
 KW pulmonary dysplasia; primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.  
 XX HOMO sapiens.  
 XX WO200186003-A2.  
 XX PD 15-NOV-2001.  
 XX P 30-JAN-2001; 2001WO-US00665.  
 XX PR 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-063256.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 04-OCT-2000; 2000US-236359P.  
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2002-114183/15.  
 XX Spatially-addressable set of single exon nucleic acid probes, used to  
 PR measure gene expression in human lung samples -  
 XX Claim 1; SEQ ID No 8607; 634pp; English.  
 XX The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridise at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labelled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label specifically bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote and (b) detecting specific hybridisation of detectably  
 CC labelled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarray having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types comprising one  
 CC of the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC

**RESULT 15**

CC ABL21524\_c  
 CC ID ABL21524 standard; DNA; 9835 BP.  
 XX AC ABL21524;  
 XX DT 26-MAR-2002 (first entry)  
 XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 16045.  
 XX XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX Drosophila melanogaster.  
 OS XX  
 PN XX  
 WO200171042-A2.  
 XX PD 27-SEP-2001.  
 XX XX  
 XX PR 23-MAR-2001; 2001WO-US0231.  
 XX PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX DR 2001-656660/75.  
 XX PA (PEKE ) PE CORP NY.  
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 CC genes from Drosophila and for elucidating cell signalling and cell-cell  
 CC interactions -  
 XX PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX PT  
 XX PR  
 XX DR  
 XX PS  
 XX Claim 1; SEQ ID NO 16045; 21PP + Sequence Listing; English.  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in development biology and in elucidating cell signalling and  
 CC

CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB5737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [ftp://wipo.int/pub/published\\_pct\\_sequences](ftp://wipo.int/pub/published_pct_sequences).  
 XX

5Q Sequence 9835 BP; 2538 A; 2196 C; 2372 G; 2729 T; 0 other;

	Query Match	Score	DB	Length
QY	3.5%	36.8	23	9835
Db	Best Local Similarity 50.0%; Matches 92; Conservative 0; Mismatches 0; Gaps 0;	Pred. No. 3.4;		
QY	CGGCCCTCGTAGGCTTCCCTAACGAATTCCCTAACACTTCCCTAGGATAAAAC	112		
Db	3019 CGGCCTCACCAAGGCTTACCAACGAACTCACAACATATTAGCTAAAGAATAGAC	2960		
QY	113 AAAGGCCTTAGAAGAATTAAAAGGTTCCTGAAACAAATACCCAGACCCGAAGCGA	172		
Db	2959 CAACGAAATAGCAATCTAAAGGTTCCTGAAACATAGAAACATAGAACACA	2900		
QY	173 AAGAGTTAAAGGGTTCTGGGATTTTTCGGCTTAAGGAGAAATTAGTCTCG	232		
Db	2899 TTTAGTAGAAAATGAAATATTATAGTATATATTATTATTATTATTTATTA	2840		
QY	233 CTAA 236			
Db	2839 GTAA 2836			

Search completed: April 15, 2003, 14:15:26  
 Job time : 1558 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 14:00:46 ; Search time 63 Seconds  
(without alignments)  
5184.298 Million cell updates/sec

Title: US-09-905-173-23  
Perfect score: 1065  
Sequence: 1 atgatacccgaggattaa.....ccctttcaagtcctctttaa 1065

Scoring table: IDENTITY NUC  
Gapod 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA: \*

```
1: /cgn2_6/podata/1/ina/5A_COMB.seq:*
2: /cgn2_6/podata/1/ina/5B_COMB.seq:*
3: /cgn2_6/podata/1/ina/6A_COMB.seq:*
4: /cgn2_6/podata/1/ina/6B_COMB.seq:*
5: /cgn2_6/podata/1/ina/PCTRUS_COMB.seq:*
6: /cgn2_6/podata/1/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1065	100.0	1065	1	US-09-999-171A-23
2	1065	100.0	1065	2	US-09-946-590B-23
3	1065	100.0	1065	3	US-09-969-326-23
4	1065	100.0	1065	4	US-09-912-184-23
C 5	61	5.7	7218	1	US-08-932-163-14
C 6	42.2	4.0	1092	2	US-08-946-590B-35
7	42.2	4.0	1092	4	US-09-412-184-35
8	34.4	3.2	12225	2	US-08-982-445-11
9	34.4	3.2	12225	4	US-09-396-540-11
10	34.4	3.2	12616	2	US-08-822-445-9
11	34.4	3.2	12616	4	US-09-396-540-9
C 12	33.2	3.1	1100	1	US-08-234-939-4
C 13	33.2	3.1	1100	1	US-08-558-865-7
C 14	33.2	3.1	1268	1	US-08-234-939-6
C 15	33.2	3.1	1268	1	US-08-558-865-6
C 16	33.2	3.1	1385	1	US-08-234-939-5
C 17	33.2	3.1	1385	1	US-08-558-865-5
C 18	33.2	3.1	1859	1	US-08-234-939-4
C 19	33.2	3.1	1859	1	US-08-558-865-4
C 20	33.2	3.1	1933	1	US-08-234-939-3
C 21	33.2	3.1	1933	1	US-08-558-865-3
C 22	33.2	3.1	3426	1	US-08-234-939-1
C 23	33.2	3.1	3426	1	US-08-558-865-1
C 24	33.2	3.1	3426	3	US-08-654-025-6
C 25	32.8	3.1	180	4	US-08-894-818B-31
C 26	32.8	3.0	1770	4	US-08-943-731-14
C 27	32.4	3.0	24183	4	US-08-943-731-3

## ALIGNMENTS

RESULT 1  
US-09-905-171A-23  
; Sequence 23, Application US/08599-171A  
; Patent No. 5814473  
; GENERAL INFORMATION:  
; APPLICANT: WARREN, Patrick V.  
; TITLE OF INVENTION: TRANSMINASES AND AMINOTRANSFERASES  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSE: CAVELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESS: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/599-171A  
; FILING DATE: Concurrently  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION: 435  
; AGENT/AGENT INFORMATION:  
; NAME: HERRON, CHARLES J.  
; REGISTRATION NUMBER: 28,019  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1065 NUCLEOTIDES  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: SINGLE  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: GENOMIC DNA  
; US-08-599-171A-23  
Query Match 100.0%; Score 1065; DB 1; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ATGATACCCAGAGGATAAGGAACCTTGAACTCACAGGGACTCTCCGCCTCC 60

US-08-646-590B-23

b 1 ATGATACCCAGGGATAAACGAACTTGAAGCGAGCTTAAGCCTTAAGAATTCGCGCTTC 60  
 b 61 GTAGGGTTTCCCTTAAGAATTCGAACTTCCCCTAACACTTCCCGAGAGATAAAACAAAGGGC 120  
 y 61 GTAGGGTTTCCCTTAAGAATTCGAACTTCCCCTAACACTTCCCGAGAGATAAAACAAAGGGC 120  
 b 121 TTAGAAGAAATTAAAAGGTTCCTCTAACGAACTTCCCAGAACAAATAACCCAGCCCCAAGGAAAGAGTTA 180  
 y 121 TTAGAAGAAATTAAAAGGTTCCTCTAACGAACTTCCCAGAACAAATAACCCAGCCCCAAGGAAAGAGTTA 180  
 b 181 AAACGGPTCTCGGATTTCGGCTTAAGGAAATTACTTCGTTACGGT 240  
 y 181 AAACGGPTCTCGGATTTCGGCTTAAGGAAATTACTTCGTTACGGT 240  
 b 241 TCGGACGAACATATACTACCTCTCAATAGCTATAAGTGTAACTTACATACCGCTTAC 300  
 y 241 TCGGACGAACATATACTACCTCTCAATAGCTATAAGTGTAACTTACATACCGCTTAC 300  
 b 301 ATACCTGTTCCACCTTCCAGTGTACAGATAAGTGTAACTTACATACCGCTTAC 360  
 y 301 ATACCTGTTCCACCTTCCAGTGTAACTTACATACCGCTTAC 360  
 b 361 GATAAGGTTAACCTGGACAAACTTGACCTTAAGCTAGATAAGTGTAACTTACATACCGCTTAC 420  
 y 361 GATAAGGTTAACCTGGACAAACTTGACCTTAAGCTAGATAAGTGTAACTTACATACCGCTTAC 420  
 b 421 GAGAAAGAAAACCCGGTACCTGGTACCTGGTACCTGGTACCTGGGAAACCTC 480  
 y 421 GAGAAAGAAAACCCGGTACCTGGTACCTGGTACCTGGGAAACCTC 480  
 b 481 TTTTCAGGGAAAGTTGAGGAGATAAGAACCCGGTACCTGGTACCTGGGAAACCTC 480  
 y 481 TTTTCAGGGAAAGTTGAGGAGATAAGAACCCGGTACCTGGTACCTGGGAAACCTC 480  
 b 541 GCCTACTATCATCTCGGAAACCTPCTCTGAAGANGCCCTCAAGGGAGATAAG 600  
 y 541 GCCTACTATCATCTCGGAAACCTPCTCTGAAGANGCCCTCAAGGGAGATAAG 600  
 b 601 CTAGTTTGGAGACACTTCAAAAATCGGTATGGGTAGGGTAGGTTATAA 660  
 y 601 CTAGTTTGGAGACACTTCAAAAATCGGTATGGGTAGGGTAGGTTATAA 660  
 b 601 GTAGTTTGGAGACACTTCAAAAATCGGTATGGGTAGGGTAGGTTATAA 660  
 y 661 GGGAAAGGGAAATTCAGAAATTAAACZAGGTGAGCTCCCTCAACGAGCTAC 720  
 y 661 GGGAAAGGGAAATTCAGAAATTAAACZAGGTGAGCTCCCTCAACGAGCTAC 720  
 b 721 CCCTCTCAAGGTATGGAAAGTCTCCCTACGGAGGGAGAAATTCTATGGAAAG 780  
 y 721 CCCTCTCAAGGTATGGAAAGTCTCCCTACGGAGGGAGAAATTCTATGGAAAG 780  
 b 781 ATAGAGGGTTGTAACAGGGAGAAAGGTAGCAGCAAATAGAAGGA 840  
 y 781 ATAGAGGGTTGTAACAGGGAGAAAGGTAGCAGCAAATAGAAGGA 840  
 b 781 ATAGAGGGTTGTAACAGGGAGAAAGGTAGCAGCAAATAGAAGGA 840  
 y 901 GAGGTTTATCAGGAGCTACTGAAAGGGATGTCCTCTCGAGGGAAACSTATCTATCGGAA 960  
 b 901 GAGGTTTATCAGGAGCTACTGAAAGGGATGTCCTCTCGAGGGAAACSTATCTATCGGAA 960  
 y 841 GTTGGGGTTTCCGGATAAGGGTAACTCTGGCTTTCAGAACGCCTAACCGGCCAC 900  
 y 841 GTTGGGGTTTCCGGATAAGGGTAACTCTGGCTTTCAGAACGCCTAACCGGCCAC 900  
 b 961 GGACTCCAAAAGTGCCTCAAGGTAAAGCTGAGGAAGAACAAAGTTCTG 1020  
 y 961 GGACTCCAAAAGTGCCTCAAGGTAAAGCTGAGGAAGAACAAAGTTCTG 1020  
 b 961 GGACTCCAAAAGTGCCTCAAGGTAAAGCTGAGGAAGAACAAAGTTCTG 1020  
 y 1021 GAAGCACTGGAGGAGGAGTATAAAATCCCTTCAGGCTCTCTCTAA 1065  
 y 1021 GAAGCACTGGAGGAGGAGTATAAAATCCCTTCAGGCTCTCTCTAA 1065  
 b 121 TTAGAAGAAATTAAAAGGTTCCTCTAACAGTATAGTGAACTTAACCGTTAACGGT 240  
 y 121 TTAGAAGAAATTAAAAGGTTCCTCTAACAGTATAGTGAACTTAACCGTTAACGGT 240  
 b 181 AAAGGGTTCTGGGAATTTCGGTTTTCGGTTAACGAAAAATTAGTCTGGTAAACGGT 240  
 y 181 AAAGGGTTCTGGGAATTTCGGTTTTCGGTTAACGAAAAATTAGTCTGGTAAACGGT 240  
 b 241 TCGGACGAACCTCATCTCATAGTATAGTGAACTTAACCGTTAACGGT 300  
 y 241 TCGGACGAACCTCATCTCATAGTATAGTGAACTTAACCGTTAACGGT 300

Sequence 23, Application US/08646590B  
 Patent No. 5,962283  
 GENERAL INFORMATION:  
 APPLICANT: Warren, Patrick V.  
 APPLICANT: Swanson, Ronald V.  
 TITLE OF INVENTION: TRANSAMINASES AND AMINO-TRANSFERASES  
 NUMBER OF SEQUENCES: 42  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish Richardson, P.C.  
 STREET: 4225 Executive Square, Suite 1400  
 CITY: La Jolla  
 STATE: CA  
 COUNTRY: US  
 ZIP: 92037  
 COMPUTER READABLE FORM:  
 COMPUTER: IBM Compatible  
 MEDIUM TYPE: Diskette  
 OPERATING SYSTEM: Windows95  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/646,590B  
 FILING DATE: 08-May-1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/599,171  
 FILING DATE: 09-FEB-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US97/01094  
 FILING DATE: 08-January-1997  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Haile, Ph.D., Lisa A.  
 REGISTRATION NUMBER: 38,347  
 REFERENCE/DOCKET NUMBER: 09010/017001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 619/678-5070  
 TELEFAX: 619/678-5099  
 INFORMATION FOR SEQ ID NO: 23:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1065 base pairs  
 LOCATION: 21-January-1997  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA  
 FEATURE:  
 NAME/KEY: Coding Sequence  
 LOCATION: 1...1062  
 US-08-646-590B-23

Query Match 100.0%; Score 1065; DB 2; Length 1065;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1055; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGATACCCAGGGATAAACGAAACTTGTAAAGGAGCTTACAAGACGGAGGTCACTCCCGCTTC 60  
 Db 1 ATGATACCCAGGGATAAACGAAACTTGTAAAGGAGCTTACAAGACGGAGGTCACTCCCGCTTC 60  
 Qy 61 GTCAGGCTTTCTCTCTAACGAAATTCCCTACGACTTTCCTGGAGGATAAACAAAGGGC 120  
 Db 61 GTCAGGCTTTCTCTCTAACGAAATTCCCTACGACTTTCCTGGAGGATAAACAAAGGGC 120  
 Qy 721 CCCTCTCAAGGTATGGAAAGTCTCCCTACGGAGGGAGAAATTCTATGGAAAG 780  
 Db 721 CCCTCTCAAGGTATGGAAAGTCTCCCTACGGAGGGAGAAATTCTATGGAAAG 780  
 Qy 781 ATAGAGGGTTGTAACAGGGAGAAAGGTAGCAGCAAATAGAAGGA 840  
 Db 781 ATAGAGGGTTGTAACAGGGAGAAAGGTAGCAGCAAATAGAAGGA 840  
 Qy 901 GAGGTTTATCAGGAGCTACTGAAAGGGATGTCCTCTCGAGGGAAACSTATCTATCGGAA 960  
 Db 901 GAGGTTTATCAGGAGCTACTGAAAGGGATGTCCTCTCGAGGGAAACSTATCTATCGGAA 960  
 Qy 841 GTTGGGGTTTCCGGATAAGGGTAACTCTGGCTTTCAGAACGCCTAACCGGCCAC 900  
 Db 841 GTTGGGGTTTCCGGATAAGGGTAACTCTGGCTTTCAGAACGCCTAACCGGCCAC 900  
 Qy 961 GGACTCCAAAAGTGCCTCAAGGTAAAGCTGAGGAAGAACAAAGTTCTG 1020  
 Db 961 GGACTCCAAAAGTGCCTCAAGGTAAAGCTGAGGAAGAACAAAGTTCTG 1020  
 Qy 1021 GAAGCACTGGAGGAGGAGTATAAAATCCCTTCAGGCTCTCTCTAA 1065  
 Db 1021 GAAGCACTGGAGGAGGAGTATAAAATCCCTTCAGGCTCTCTCTAA 1065

QY 301 ATACCTGTCACCCATTCCATGTCAGAGATAAGTGGAAAGTTCTGGAAAGACCCCTC 360  
Db 301 ATACCTGTCACCCATTCCATGTCAGAGATAAGTGGAAAGTTCTGGAAAGACCCCTC 360  
Qy 361 GTAARGGTCACTGGACAAAACCTTGATAGACTTAGAAGAAAGTATGAAATTATAA 420  
Db 361 GTAAAGGTCACTGGACAAAACCTTGATAGACTTAGAAGAAAGTATGAAATTATAA 420  
Qy 421 GAGAAAGAAAACCGTTCGCGTACTTGTACCCAACAACCCCACGGAAACCTC 480  
Db 421 GAGAAAGAAAACCGTTCGCGTACTTGTACCCAACAACCCCACGGAAACCTC 480  
Qy 481 TTTCAGGGAAAGATTCAGGAGATAAGAACGGGTGTTCTGTAAAGCAGAA 540  
Db 481 TTTCAGGGAAAGATTCAGGAGATAAGAACGGGTGTTCTGTAAAGCAGAA 540  
Qy 541 GCCTACTATCAACTCGGAAACCTTCTGGAGACGCCCTAAAGGGAGATAAGC 600  
Db 541 GCCTACTATCAACTCGGAAACCTTCTGGAGACGCCCTAAAGGGAGATAAGC 600  
Qy 601 GTAGTTTGGGAAAGTTAGGAGATAAGAACAGGGTGTITTCAGTAAAGCAGAA 660  
Db 601 GTAGTTTGGGAAAGTTAGGAGATAAGAACAGGGTGTITTCAGTAAAGCAGAA 660  
Qy 661 GGGAAAGGGAAATCTCAGRAATTAAACAGGTGAGACCCCCCTCAAGTGACCTAC 720  
Db 661 GGGAGGGGAATCTCAGRAATTAAACAGGTGAGACCCCCCTCAAGTGACCTAC 720  
Qy 721 CCCTCTCAGGTGATGCCAAAGTTCTCCTACGGGGAGAGAAATTCTATGAAAG 780  
Db 721 CCCTCTCAGGTGATGCCAAAGCTCTCCTACGGGGAGAGAAATTCTATGAAAG 780  
Qy 781 ATACGGGTTGAGGAAAGTTCTCAGGGAGAAAGTGTAGAAATTACAGGTGAGAAATAGAGGA 840  
Db 781 ATACGGGTTGAGGAAAGTTCTCAGGGAGAAAGTGTAGAAATTACAGGTGAGAAATAGAGGA 840  
Qy 841 GTGAGGTTTCGAGTAAGCTCAACTCTGCTTCAAGAACGCCCTAACCCGCCAC 900  
Db 841 GTGAGGTTTCGAGTAAGCTCAACTCTGCTTCAAGAACGCCCTAACCCGCCAC 900  
Qy 901 GAGGTATCAGGAGTACTGAAAGGTATGTCCTGTCAGGAACGTATCTACATGGAA 960  
Db 901 GAGGTATCAGGAGTACTGAAAGGTATGTCCTGTCAGGAACGTATCTACATGGAA 960  
Qy 961 GGACTCCAAAGTGCCTCAGGTTAACGGAAACACAGTTCTCG 1020  
Db 961 GGACTCCAAAGTGCCTCAGGTTAACGGAAACACAGTTCTCG 1020  
Qy 1021 GAAGCACTGAGGAGTATAAACCTCTCAAGCTCTTTAA 1065  
Db 1021 GAAGCACTGAGGAGTATAAACCTCTCAAGCTCTTTAA 1065

## RESULT 3

US-09-069-226-23

; Sequence 23, Application US/09069226

; Patent No. 6013509

## GENERAL INFORMATION:

; APPLICANT: WARREN, Patrick V.

; TITLE OF INVENTION: TRANSAMINASES AND AMINO TRANSFERSSES

; NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESS: CECCHI, STEWART & OLSTEIN

; STREET: 6 BECKER FARM ROAD

; CITY: ROSELAND

; STATE: NEW JERSEY

; COUNTRY: USA

; ZIP: 07068

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2

; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/069, 226  
; FILING DATE:  
; CLASSIFICATION:  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 08/599, 171  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HERRON, CHARLES J.  
; REGISTRATION NUMBER: 28, 019  
; REFERENCE/DOCKET NUMBER: 331400-38  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INQUIRY FOR SPO ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1065 NUCLEOTIDES  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: SINGLE  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: GENOMIC DNA  
; US-09-069-226-23

Query Match 100.0%; Score 1065; DB 3; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGATACCCAGAGGAGTTAGGAAACTTGAAAGCTTACAAAGCAGGGAGTCACTCCGGCCTTC 60  
Db 1 ATGATACCCAGAGGAGTTAGGAACTTGAAAGCTTACAAAGCAGGGAGTCACTCCGGCCTTC 60  
Qy 61 GTCAAGCTTTCCTCTAAGCAAATCCCTACTGACTTTCCCTACTGAAATTTCCCTGACAATACCCAC 120  
Db 61 GTCAAGCTTTCCTCTAAGCAAATCCCTACTGACTTTCCCTACTGAAATTTCCCTGACAATACCCAC 120  
Qy 121 TTAGAGAATTAAAAGGTTTCCCTGACAATACCCAC 180  
Db 121 TTAGAGAATTAAAAGGTTTCCCTGACAATACCCAC 180  
Qy 181 AAAGGGTCTTGGGATTTTGGGTTAAGGAAGAAATTAGTTCTGGTAAAGGT 240  
Db 181 AAAGGGTCTTGGGATTTTGGGTTAAGGAAGAAATTAGTTCTGGTAAAGGT 240  
Qy 241 TCGGAGCACTCATACCTACCTTCAAGTGTAACTTCAATACCCGTITAC 300  
Db 241 TCGGAGCACTCATACCTACCTTCAAGTGTAACTTCAATACCCGTITAC 300  
Qy 301 ATACCTGTCACCTTCCATGAGATAAGTGGAAAGACCCCTC 360  
Db 301 ATACCTGTCACCTTCCATGAGATAAGTGGAAAGACCCCTC 360  
Qy 361 GTAAAGGTCACTGGAGAAACTTGTATGACTTAAAGAAGTATTGATTATAA 420  
Db 361 GTAAAGGTCACTGGAGAAACTTGTATGACTTAAAGAAGTATTGATTATAA 420  
Qy 421 GAGAAAGAAAACCGTTCGCGTACTTGTCTTACCCACGGAAACCTC 480  
Db 421 GAGAAAGAAAACCGTTCGCGTACTTGTCTTACCCACGGAAACCTC 480  
Qy 481 TTTCAGGGAAAGATTCAGGAGATAAGAAACAGGGTTCTGTAAATAGCAGAA 540  
Db 481 TTTCAGGGAAAGATTCAGGAGATAAGAAACAGGGTTCTGTAAATAGCAGAA 540  
Qy 541 GCCTACTATCATCTACTCGGAGAAACCTTCTCGGACTCTCGGTACTTGTCTACCAACACCCACGGAAACCTC 600  
Db 541 GCCTACTATCATCTACTCGGAGAAACCTTCTCGGACTCTCGGTACTTGTCTACCAACACCCACGGAAACCTC 600  
Qy 601 GTACTTGTAGGAGACTTCAAAATCGTATGGGTTAAGGTAGGATTATAA 660  
Db 601 GTACTTGTAGGAGACTTCAAAATCGTATGGGTTAAGGTAGGATTATAA 660

TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA  
 FEATURE:  
 NAME/KEY: Coding Sequence  
 LOCATION: 1...1062  
 US-09-412-184-23

Query Match 100.0%; Score 1065; DB 4; Length 1065;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1065; Conservative 0; Mismatches 0;  
 Indels 0; Gaps 0;

1 ATGATACCCAGAGGATTAAAGGAACCTTGAGCTTACAAGCAGGGAGGTCACTCCGCCAC 60  
 1 ATGATACCCAGAGGATTAAAGGAACCTTGAGCTTACAAGCAGGGAGGTCACTCCGCCAC 60

61 GTCAGGCTTCCTCTAACAAATCCCCATTAACGAAATTCGGAGGATAAAACAAAGGCC 120  
 61 GTCAGGCTTCCTCTAACAAATCCCCATTAACGAAATTCGGAGGATAAAACAAAGGCC 120

621 TTAGAGAATTAAAAGGTTCCCTGACAAATACCCAGACCCCAGAAACAGAAAGAGTTA 180  
 621 TTAGAGAATTAAAAGGTTCCCTGACAAATACCCAGACCCCAGAAACAGAAAGAGTTA 180

901 GAGGTATCAGGAGTACTGAAAGGATGCTCTGTCAGGAAGTATCTTACATGGAA 960  
 901 GAGGTATCAGGAGTACTGAAAGGATGCTCTGTCAGGAACGCTATCTTACATGGAA 960

961 GGAACCAAATGGCCCTAGGGTAAGGTAGGGAAACCGGAGAAACACAAAGPTCTG 1020  
 961 GGAACCAAATGGCCCTAGGGTAAGGTAGGGAAACCGGAGAAACACAAAGPTCTG 1020

1021 GAAAGCTTAAAGGAGACTATAATCCCTTCAAGCTCTTTAA 1065  
 1021 GAAAGCTTAAAGGAGACTATAATCCCTTCAAGCTCTTTAA 1065

**RESULT 4**  
 US-09-412-184-23  
 Sequence 23, Application US/09412184  
 Patent No. 6268188

GENERAL INFORMATION:  
 APPLICANT: Warren, Patrick V.  
 APPLICANT: Swanson, Ronald V.  
 TITLE OF INVENTION: TRANSMINASES AND AMINOTRANSFERASES  
 NUMBER OF SEQUENCES: 42  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Richardson, P.C.  
 STREET: 4225 Executive Square, Suite 1400  
 CITY: La Jolla  
 STATE: CA  
 COUNTRY: US  
 ZIP: 92037  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER SYSTEM: Windows 95  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/412,184  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/646,590  
 FILING DATE: 08-MAY-1996  
 APPLICATION NUMBER: 08/599,171  
 FILING DATE: 09-FEB-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US97/01094  
 FILING DATE: 21-January-1997  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Haile, Ph.D., Lisa A.  
 REGISTRATION NUMBER: 38,347  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 619/678-5070  
 TELEFAX: 619/678-5099  
 INFORMATION FOR SEQ ID NO: 23:  
 LENGTH: 1065 base pairs

1 GGAAGGGAAATTCGGAGGAAACCTTGCTTACCCGGCTAACCTCTGCTTACCCGCCAC 900  
 1 GTTGAGGTTTCGGAGTAAAGGAGAAATTAACAGGTCACTCCCTCACGTGCTTACCCGCCAC 900





NAME/KEY: CDS  
; LOCATION: 190..11208  
Query Match 3 : 28; Score 34.4; DB 2; Length 12225;  
Best Local Similarity 48.0%; Pred. No. 1.6;  
Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;  
Qy 102 GGAGATAAACAAAGGCCCTAGAGAAATTAAAAAGGTTCCTGAAACAAATACCCAGA 161  
Db 8430 GGATCAAAATCAAGGATAATATCTAAATAGCAGTATACCCAGGCAGTGTCCT 8489  
Qy 162 CCCGAAGGGAAAGAGTTAAAGGGTTCCTGGGATTTTGGCGTYYAGGAAGAAA 221  
Db 8490 CTCCCAAGGAAATGAGAAATGAGGAAAGGTGTTAAAGCTAGA 8549  
Qy 222 TTAGTTCTCGGTAACTACGGTTCGGAGACTCATATACCTCTCAATAGTATAGGTA 281  
Db 8550 TTGAGTGCAGCACATGGCAGAACCTTCAGGAGCTACATGAGCAGT 8609  
Qy 282 ACTTACATACCCGTTACATACCC 305  
Db 8610 ATGGTATGCCCATACTATCC 8633

RESULT 9  
US-09-396-540-11  
; Sequence 11, Application US/09396540  
; Patient No. 6310102  
; GENERAL INFORMATION:  
; APPLICANT: Kaplan, Jerry  
; APPLICANT: Perou, Charles  
; APPLICANT: Moore, Karen  
; TITLE OF INVENTION: COMPOSITIONS FOR THE DIAGNOSIS  
; AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036/2711  
; COMPUTER READABLE FORM:  
; COMPUTER TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/396,540  
; FILING DATE: 08/09/2000  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/822,445  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Corruzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-062-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-669-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12225 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 190..11208

Query Match 3 : 28; Score 34.4; DB 2; Length 12225;  
Best Local Similarity 48.0%; Pred. No. 1.6;  
Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

Query Match 3 : 28; Score 34.4; DB 2; Length 12616;  
Best Local Similarity 48.0%; Pred. No. 1.6;  
Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

RESULT 10  
US-08-822-445-9  
; Sequence 9, Application US/08822445  
; Patent No. 595223  
; GENERAL INFORMATION:  
; APPLICANT: Kaplan, Jerry  
; APPLICANT: Perou, Charles  
; APPLICANT: Moore, Karen  
; TITLE OF INVENTION: COMPOSITIONS FOR THE DIAGNOSIS  
; AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036/2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/822,445  
; FILING DATE: 21-MAR-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Corruzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-062-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12616 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 190..11592

Query Match 3 : 28; Score 34.4; DB 2; Length 12616;  
Best Local Similarity 48.0%; Pred. No. 1.6;  
Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

RESULT 11  
 QY 102 GGAGATAAAACAAAGGCCCTTAAAGAATTAAAAAGGTTCCCTTGAAACAAATACCCAGA 161  
 Db 8817 GGATTCAAAATCAAAGGATATATCTAAATAATAGCTGCAGATATCACCCAGCAGTGTCTT 8876  
 ; Sequence 9, Application US/09396540  
 ; Patent No. 6310182  
 GENERAL INFORMATION:  
 APPLICANT: Kaplan, Jerry  
 APPLICANT: Perou, Charles  
 APPLICANT: Moore, Karen  
 TITLE OF INVENTION: COMPOSITIONS FOR THE DIAGNOSIS  
 TITLE OF INVENTION: AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME  
 NUMBER OF SEQUENCES: 32  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Penne & Edmonds  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036/2711  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/396,540  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/822,445  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Cotuzzi, Laura A.  
 REGISTRATION NUMBER: 30,742  
 REFERENCE/DOCKET NUMBER: 7853-062-999  
 TELEPHONE: 212-790-9090  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 12616 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 190..11592  
 US-09-396-540-9  
 Query Match 3.28; Score 34.4; DB 4; Length 12616;  
 Best Local Similarity 4.0%; Pred. No. 1..6; Indels 0; Gaps 0;  
 Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

RESULT 12  
 US-08-234-939-7/c  
 ; Sequence 7, Application US/08234939  
 ; Patent No. 5459352  
 GENERAL INFORMATION:  
 APPLICANT: Conkling, Mark A.  
 APPLICANT: Yamamoto, Yuri T.  
 TITLE OF INVENTION: Root Specific Gene Promoter  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Kenneth D. Sibley; Bell, Seltzer, Park and Gibson  
 STREET: Post Office Drawer 34009  
 CITY: Charlotte  
 STATE: No. 545252th Carolina  
 COUNTRY: U.S.A.  
 ZIP: 28234  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.24  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/234,939  
 FILING DATE:  
 CLASSIFICATION: 800  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US/07/649,564  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Sibley, Kenneth D.  
 REGISTRATION NUMBER: 31,665  
 REFERENCE/DOCKET NUMBER: 5051-141  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-881-3140  
 TELEFAX: 919-881-3175  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1100 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-234-939-7

Query Match 3.18; Score 33.2; DB 1; Length 1100;  
 Best Local Similarity 51.3%; Pred. No. 0.88; Mismatches 73; Indels 0; Gaps 0;  
 Matches 77; Conservative 0;

QY 267 AATAGCTTAGGGTAACTTAACATCCCTGTCACCTTCAGCATGTA 326  
 Db 331 ATTTGTTAAGATGTCCTCACATCCCTTTAGAACTATTCACTTTGATGTC 272  
 QY 327 CGAGATAAGTGGAAGTCTGGAAAGACCCCTCGTAAAGGTCAACNGACGAAACT 386  
 Db 271 TTGAGTTGTGAATTTCCTAAAGGAAAATAGTTGCATCAAACGATTAAANGTA 212  
 QY 102 GGAGATAAAACAAAGGCCCTTAAAGAATTAAAAAGGTTCCCTTGAAACAAATACCCAGA 161

QY 387 TGATATAGACTTAGAAAAGAAGTATTGAAATT 416  
 Db 211 TGCCTAAATTCAAAATTTGGGGTT 182

RESULT 13  
 US-08-558-865-7/c  
 Sequence 7, Application US/08558865  
 Patent No. 5750386

GENERAL INFORMATION:  
 APPLICANT: Conkling, Mark A.  
 APPLICANT: Opperman, Charles H.  
 APPLICANT: Taylor, Christopher G.  
 TITLE OF INVENTION: Root Specific Gene Promoter  
 NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Kenneth D. Sibley; Bell, Seltzer, Park and Gibson  
 STREET: Post Office Drawer 34009  
 CITY: Charlotte  
 STATE: No. 5750386th Carolina  
 COUNTRY: U.S.A.  
 ZIP: 28234

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 FILING DATE:  
 APPLICATION NUMBER: US/08/558, 865

CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 FILING DATE:  
 APPLICATION NUMBER: US/08/236, 678

CLASSIFICATION: 800  
 APPLICATION NUMBER: US/07/770, 082

FILING DATE:  
 APPLICATION NUMBER: 800  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 FILING DATE:  
 APPLICATION NUMBER: US/08/236, 678

CLASSIFICATION: 800  
 APPLICATION NUMBER: US/07/770, 082

FILING DATE:  
 NAME: Sibley, Kenneth D.  
 REGISTRATION NUMBER: 31,665  
 ATTORNEY/AGENT INFORMATION:  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-881-3140  
 TELEFAX: 919-881-3175  
 TELEX: 575102

INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1100 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)

QY 267 AATAGCTTATAGTGAACTTTACATACCCCTTACATACTGTGAAAGTCTGCCACCTTCCCAGTA 326  
 Db 331 ATTTGATTAAGATGTCCTCACATCCCTTACAACATTTCTACTTTTGATTGTC 272

QY 267 AATAGCTTATAGTGAACTTTACATACCCCTTACATACTGTGAAAGTCTGCCACCTTCCCAGTA 326  
 Db 331 ATTTGATTAAGATGTCCTCACATCCCTTACAACATTTCTACTTTTGATTGTC 272

QY 327 CGAGATAAGTGGAAAAGTCTGGAAAGCCCTCGTAAGGTTCAACTGGACAAACTT 386  
 Db 439 TTAGATGTGTGATAATTCTCAAAGAAAATAGTGTCAAAACCATCTTAAAGATA 380

QY 387 TGATATAGACTTAGAAAAGAAGTATTGAAATT 416  
 Db 211 TGCCTAAATTCAAAATTTGGGGTT 182

RESULT 14

ADDRESSEE: Gibson  
 STREET: Post Office Drawer 34009  
 CITY: Charlotte  
 STATE: No. 5750386th Carolina  
 COUNTRY: U.S.A.  
 ZIP: 28234  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/558,865  
 FILING DATE:  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/236,678  
 FILING DATE:  
 CLASSIFICATION: 800  
 APPLICATION NUMBER: US/07/770,082  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Sibley, Kenneth D.  
 REGISTRATION NUMBER: 31,665  
 REFERENCE/DOCKET NUMBER: 5051-166  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-881-3140  
 TELEFAX: 919-881-3175  
 TELEX: 575102  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1268 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-558-865-6

Query Match 3.1%; Score 33.2; DB 1; Length 1268;  
 Best Local Similarity 51.3%; Pred. No. 0.96;  
 Matches 77; Conservatives 0; Mismatches 73; Indels 0; Gaps 0;  
 QY 267 ATAGCTATAGTGAACTTACATACCGGTACATACCTGTCACCCACCTTCCCATATGTA 326  
 Db 499 ATTGGATTAAGATGCCCCAACGCCCTTCAACTATTTCTACTTTTGATGTGTC 440  
 QY 327 CGAGATACTGGAAAGTTCTCGGAAGACCCCTCGTAAAGGTCAACTGGACGAAACTT 386  
 Db 439 TTAGATGTTGTGATATTCTCAAAGGAAATAGTTCATAAACGCATCTAAAGTA 380  
 QY 387 TGTATAGACTAGAAAGAATGAAATTGAATT 416  
 Db 379 TGCCATAAATTCAAATTTGGTT 350

Search completed: April 15, 2003, 15:49:10  
 Job time : 100 secs

GenCore version 5.1.3  
 Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 15:15:16 ; Search time 125 Seconds

(without alignments)  
 7473.469 Million cell updates/secTitle: US-09-905-173-23  
 Perfect score: 1065

Sequence: atgatacccaaggattaa.....cccttcaaggctctttaa 1065

Scoring table: IDENTITY NUC  
 Gapext 1.0  
 Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 438583890 residues

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
 Maximum Match 100%

Listing first 45 summaries

Database :

```
Published_Applications_NA:
1: /cgn2_6/ptodata/2/pubpna/us07_purcomb.seq;*
2: /cgn2_6/ptodata/2/pubpna/pct_new_pub.seq;*
3: /cgn2_6/ptodata/2/pubpna/us06_new_pub.seq;*
4: /cgn2_6/ptodata/2/pubpna/us06_purcomb.seq;*
5: /cgn2_6/ptodata/2/pubpna/us07_new_pub.seq;*
6: /cgn2_6/ptodata/2/pubpna/pctus_pubcomb.seq;*
7: /cgn2_6/ptodata/2/pubpna/us08_new_pub.seq;*
8: /cgn2_6/ptodata/2/pubpna/us08_pubcomb.seq;*
9: /cgn2_6/ptodata/2/pubpna/us09_new_pub.seq;*
10: /cgn2_6/ptodata/2/pubpna/us09_pubcomb.seq;*
11: /cgn2_6/ptodata/2/pubpna/us10_new_pub.seq;*
12: /cgn2_6/ptodata/2/pubpna/us10_pubcomb.seq;*
13: /cgn2_6/ptodata/2/pubpna/us60_new_pub.seq;*
14: /cgn2_6/ptodata/2/pubpna/us60_pubcomb.seq;*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	1065	100.0	1065	9	US-10-060-432-23		Sequence 23, Appli
2	1065	100.0	1065	10	US-09-905-173-23		Sequence 23, Appli
3	46.2	4.3	1085	10	US-09-974-300-298		Sequence 298, App
4	42.2	4.0	1092	10	US-09-905-173-35		Sequence 35, Appli
5	37.8	3.5	198	9	US-10-083-357-579		Sequence 579, App
6	37.6	3.5	588	10	US-09-864-761-6646		Sequence 6646, App
7	36.2	3.4	6804	7	US-08-781-986A-365		Sequence 365, App
8	35.8	3.4	560	10	US-09-974-300-4716		Sequence 4716, App
9	35.8	3.4	1098	9	US-09-738-626-2298		Sequence 2298, App
10	35.2	3.3	1350	9	US-09-971-536-8		Sequence 8, Appli
c 11	34.6	3.2	2000	9	US-09-938-842A-3936		Sequence 336, Ap
c 12	34.4	3.2	375	10	US-09-864-761-1114		Sequence 11114, A
c 13	34.4	3.2	2270	10	US-09-887-576-455		Sequence 455, App
c 14	34.4	3.2	12225	10	US-09-927-668-11		Sequence 11, Appli
c 15	34.4	3.2	12616	10	US-09-927-668-9		Sequence 9, Appli
c 16	34.2	3.2	534	9	US-09-738-626-2011		Sequence 2011, Appli
c 17	34.2	3.2	1093	9	US-10-001-876-78		Sequence 78, Appli
c 18	34.2	3.2	1691139	9	US-10-067-514-1		Sequence 1, Appli
c 19	34.2	3.2	3309400	9	US-09-738-626-8		Sequence 5128, Ap

## ALIGNMENTS

```
RESULT 1
US-10-060-432-23
; Sequence 23, Application US-10060432
; Publication No. US2004092a1
; GENERAL INFORMATION:
; APPLICANT: WARREN, Patrick V.
; INVENTION: TRANSINASES AND AMINOTRANSFERASES
; TITLE OF INVENTION: TRANSINASES AND AMINOTRANSFERASES
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US-10-060-432-2
; FILING DATE: 29-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 09/481,733
; FILING DATE: 11-JAN-2000
; APPLICATION NUMBER: US-08/599,171
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION
; NAME: HERRON, CHARLES J.
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-38
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1065 NUCLEOTIDES
; TYPE: NUCLEIC ACID
```

SEQUENCE DESCRIPTION: SEQ ID NO: 23:	
STRANDEDNESS: SINGLE	
TOPOLOGY: LINEAR	
MOLECULE TYPE: GENOMIC DNA	
SEQUENCE DESCRIPTION: SEQ ID NO: 23:	-10-060-432-23
Query Match	100.0% Score 1065; DB 9; Length 1065;
Best Local Similarity	100.0% Pred. No. 0;
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
1 ATGATACCAGAGGATAAGGAACCTGGACTTCAAGACGGAGGTACTCCGCCCTC 60	
1 ATGATACCAGAGGATAAGGAACCTGGACTTCAAGACGGAGGTACTCCGCCCTC 60	
61 GTCAAGCTTTCCTCTAACGAAATTCCCTAGACTTTCGGAGGATAAACAAAGGCC 120	
61 GTCAAGCTTTCCTCTAACGAAATTCCCTAGACTTTCGGAGGATAAACAAAGGCC 120	
121 TTAGAAGAATTAAAAGGTTCCCTTGAAACAAATTCCAGACCCGAAAGAGTA 180	
121 TTAGAAGAATTAAAAGGTTCCCTTGAAACAAATTCCAGACCCGAAAGAGTA 180	
181 AAAGCGGTTCTGGGATTTTCGGCTTAAGGAAGAAATTAGTCTCGSTAACGGT 240	
181 AAAGCGGTTCTGGGATTTTCGGCTTAAGGAAGAAATTAGTCTCGSTAACGGT 240	
241 TCGGAGGAACCTCATATACCTCTCATAGCTTACAGTCAATACCGGTTAC 300	
241 TCGGAGGAACCTCATATACCTCTCATAGCTTACAGTCAATACCGGTTAC 300	
301 ATACCTGTTCCACCTTCCATGTAAGTGGAAAGACTCTVCGGAAGGCCCTC 360	
301 ATACCTGTTCCACCTTCCATGTAAGTGGAAAGACTCTVCGGAAGGCCCTC 360	
361 GTAAAGGTTCACTGGAAACCTTGATAGACTTAGAAGATATTGAAATTAAATA 420	
361 GTAAAGGTTCACTGGAAACCTTGATAGACTTAGAAGATATTGAAATTAAATA 420	
421 GAGAAAGAAAACCCGTTCGGGTTACCTTACCAAAACCCACGGGAAACCTC 480	
421 GAGAAAGAAAACCCGTTCGGGTTACCTTACCAAAACCCACGGGAAACCTC 480	
481 TTTTCCAGGGAAAGATGGAGGATAAGAACAGGGGTGTTCTGTGTAATAGACGAA 540	
481 TTTTCCAGGGAAAGATGGAGGATAAGAACAGGGGTGTTCTGTGTAATAGACGAA 540	
541 GCCTACTATCATTACTCGGGAAACCTTCTGGAAACCGGCACAAAGGAAGATACG 600	
541 GCCTACTATCATTACTCGGGAAACCTTCTGGAAACCGGCACAAAGGAAGATACG 600	
601 GTAGTTTGAGGACACTTCAGGAAAGTGGATGAGGATAACAGGTGACCTAC 720	
601 GTAGTTTGAGGACACTTCAGGAAAGTGGATGAGGATAACAGGTGACCTAC 720	
721 CCTCTCTAGGTATGGAAAGTCTCTACGGAGGAAAGTAATGGAAAG 780	
721 CCTCTCTAGGTATGGAAAGTCTCTACGGAGGAAAGTAATGGAAAG 780	
781 ATACAGGAGGTAAAGGCTCAAGGGAAAGTGGATGAGGATAATAGAGGA 840	
781 ATACAGGAGGTAAAGGCTCAAGGGAAAGTGGATGAGGATAATAGAGGA 840	
841 GTTGAGGTTTCGGACTTAAGGCTAACTTCGTTCAAGACGCCCTACGCCAC 900	
841 GTTGAGGTTTCGGACTTAAGGCTAACTTCGTTCAAGACGCCCTACGCCAC 900	
901 GAGGTTATCAGGAGCTACTGAAAAGGATGCTGAACTCATAGGAA 960	
901 GAGGTTATCAGGAGCTACTGAAAAGGATGCTGAACTCATAGGAA 960	
QY 961 GGACTCCAAAAGTGCTCAGGGTAGGGAAACCGGAACAAACAGTTCTG 1020	
Db 961 GGACTCCAAAAGTGCTCAGGGTAGGGAAACCGGAACAAACAGTTCTG 1020	
QY 1021 GAACCACTGGAGGAGATAAAATCCCTTCAGCTCTCTATA 1065	
Db 1021 GAACCACTGGAGGAGATAAAATCCCTTCAGCTCTCTATA 1065	
RESULT 2	
US-09-905-173-23	
Sequence 23; Application US/09905173	
; Patent No. US20020132295A1	
; GENERAL INFORMATION:	
; APPLICANT: DIVERSA CORPORATION	
; SHORT, JAY M.	
; APPLICANT: WARREN, PATRICK V.	
; APPLICANT: SWANSON, RONALD V.	
; APPLICANT: MATIUR, ERIC J.	
; TITLE OF INVENTION: ENZYMES HAVING TRANSAMINASE AND AMINOTRANSFERASE ACTIVITY AND	
; TITLE OF INVENTION: USE THEREOF	
; FILE REFERENCE: DIVER1240-7	
; CURRENT APPLICATION NUMBER: US/09/905-173	
; CURRENT FILING DATE: 2001-07-12	
; PRIOR APPLICATION NUMBER: US 09/412,184	
; PRIOR FILING DATE: 1999-10-04	
; PRIOR APPLICATION NUMBER: US 09/389,537	
; PRIOR FILING DATE: 1999-09-02	
; PRIOR APPLICATION NUMBER: US 09/646,590	
; PRIOR FILING DATE: 1996-05-08	
; PRIOR APPLICATION NUMBER: US 08/599,171	
; PRIOR FILING DATE: 1996-02-09	
; PRIOR APPLICATION NUMBER: US 09/481,733	
; PRIOR FILING DATE: 2000-01-11	
; PRIOR APPLICATION NUMBER: US 09/069,226	
; PRIOR FILING DATE: 1998-04-27	
; NUMBER OF SEQ ID NOS: 40	
; SEQ ID NO: 23	
; LENGTH: 1065	
; TYPE: DNA	
; ORGANISM: Aquifex	
US-09-905-173-23	
Query Match	100.0%; Score 1065; DB 10; Length 1065;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1 ATGAAACCCGAGATTAGGACTTCCCTTAACGAAATTCCCTAGACTCCGCCCTC 60	
Db 1 ATGAAACCCGAGATTAGGACTTCCCTTAACGAAATTCCCTAGACTCCGCCCTC 60	
QY 61 GTCAAGGTTTCCTCTTAACGAAATTCCCTAGACTTCCGCCCTCAGCTTCCGCCCTC 60	
Db 61 GTCAAGGTTTCCTCTTAACGAAATTCCCTAGACTTCCGCCCTCAGCTTCCGCCCTC 60	
QY 121 TTAGAAAGATAAAAGGTTCCCTTAACGAAATTCCCTAGACTTCCGCCCTCAGCTTCCGCCCTC 120	
Db 121 TTAGAAAGATAAAAGGTTCCCTTAACGAAATTCCCTAGACTTCCGCCCTCAGCTTCCGCCCTC 120	
QY 181 AAAGCGGTTTCCTGCGGATTTTCGCGTTAAGGAGAAATTAGTTCGGTACGGT 240	
Db 181 AAAGCGGTTTCCTGCGGATTTTCGCGTTAAGGAGAAATTAGTTCGGTACGGT 240	
QY 241 TCGGAGGAACCTCATATACTACCTCTCAATAGCTTAACATCCGTTAC 300	
Db 241 TCGGAGGAACCTCATATACTACCTCTCAATAGCTTAACATCCGTTAC 300	
QY 301 ATACCTGTTCCACCTTCCATGAGATAAGTTCGGAAAGACCCCTC 360	
Db 301 ATACCTGTTCCACCTTCCATGAGATAAGTTCGGAAAGACCCCTC 360	

Query Match Score 46.2; DB 10; Length 1085;  
 Best Local Similarity 47.4%; Pred. No. 0.00077;  
 Matches 138; Conservative 0; Mismatches 153;  
 Indels 0; Gaps 0;

QY 361 GTAAAGGTCAACTGGACGAAACCTTGATATAGACTTAAAGAGTATTGAATTATA 420  
 Db 361 GTAAAGGTCAACTGGACGAAACCTTGATATAGACTTAAAGAGTATTGAATTATA 420

QY 421 GAGAAAGAAAACCGTTCTGGTACTTGCTTACCAAAACCCACGGAAACCTC 480  
 Db 421 GAGAAAGAAAACCGTTCTGGPACTTGCTTACCCAAACACCCACGGAAACCTC 480

QY 481 1TTTCCAGGGAAAGATTGAGGATAAGAAACAGGGTTCTGTTAATAGACGAA 540  
 Db 481 1TTTCCAGGGAAAGATTGAGGATAAGAAACAGGGTTCTGTTAATAGACGAA 540

QY 541 GCCTACTATCATTAATCCTGGAAACCTTCTGGAAACGCCTCAAAGGGAAAGATAG 600  
 Db 541 GCCTACTATCATTCCTGGAAACCTTCTGGAAACGCCTCAAAGGGAAAGATAG 600

QY 601 GTAGTTTGGAGGACACTTCAAATCGGTATGGGAGTTAAAGGTAGGGATTATA 660  
 Db 601 GTAGTTTGGAGGACACTTCAAATCGGTATGGGAGCTTAAAGGTAGGGATTATA 660

QY 661 GGAAGGGAAATCTCTCAGAAATTAAACAGGTGAGACTCCCTCACGTGACCTAC 720  
 Db 661 GGAAGGGAAATCTCTCAGAAATTAAACAGGTGAGACTCCCTCACGTGACCTAC 720

QY 721 CCCTCTCAGGTATGCCAAAGTCTCCATCGGAGGAGAAATTCCCTAATGGAAAAAG 780  
 Db 721 CCCTCTCAGGTATGCCAAAGTCTCCATCGGAGGAGAAATTCCCTAATGGAAAAAG 780

RESULT 4  
 US-09-905-173-35

; Sequence 35, Application US/09905173  
 ; Patent No. US2002013295A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DIVERSA CORPORATION  
 ; ATTORNEY: SHORT, JAY M.  
 ; APPLICANT: WARREN, PATRICK V.  
 ; APPLICANT: SWANSON, RONALD V.  
 ; APPLICANT: MATHUR, ERIC J.  
 ; TITLE OF INVENTION: ENZYMES HAVING TRANSAMINASE AND AMINOTRANSFERASE ACTIVITY AND  
 ; TITLE OF INVENTION: USE THEREOF  
 ; CURRENT APPLICATION NUMBER: US/09/905.173  
 ; CURRENT FILING DATE: 2001-07-12  
 ; PRIOR APPLICATION NUMBER: US 09/412,184  
 ; PRIOR FILING DATE: 1999-10-04  
 ; PRIOR APPLICATION NUMBER: US 09/389,537  
 ; PRIOR FILING DATE: 1999-09-02  
 ; PRIOR APPLICATION NUMBER: DIVER1240-7  
 ; FILE REFERENCE: DIVER1240-7  
 ; PRIOR FILING DATE: 1996-05-08  
 ; PRIOR APPLICATION NUMBER: US 08/599,171  
 ; PRIOR FILING DATE: 1996-02-09  
 ; PRIOR APPLICATION NUMBER: US 09/481,733  
 ; PRIOR FILING DATE: 2000-01-11  
 ; PRIOR APPLICATION NUMBER: US 09/063,226  
 ; PRIOR FILING DATE: 1998-04-27  
 ; NUMBER OF SEQ ID NOS: 40  
 ; SOFTWARE: Patent version 3.0  
 ; SEQ ID NO: 35  
 ; LENGTH: 1092  
 ; TYPE: DNA  
 ; ORGANISM: Ammonifex degensii  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (986). (987)  
 ; OTHER INFORMATION: n is any nucleotide  
 US-09-905-173-35

Query Match Score 4.0%; DB 10; Length 1092;  
 Best Local Similarity 48.2%; Pred. No. 0.013;  
 Matches 119; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 138 GGTTCCCTGAAATAACCCAGAACCCGAAGGAGTAAAGCGGTTCTGGGA 197  
 Db 168 GGACCACTGGCACCTTACCCAGAAAGCTCAAGCTATGACGGCTGGTAA 227

QY 198 TTTTTGGCTTAAGGAAGAAAATTAGTTCTGGTAACGGTTACGGTAACCTATA 257

RESULT 3  
 US-09-974-300-298

; Sequence 298, Application US/09974300  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Berka, Randy M.  
 ; APPLICANT: Clausen, Ib Groth  
 ; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
 ; TITLE OF INVENTION: Expression  
 ; FILE REFERENCE: 10085\_500-US  
 ; CURRENT APPLICATION NUMBER: US/09/974,300  
 ; PRIOR APPLICATION NUMBER: 09/680,598  
 ; PRIOR FILING DATE: 2000-10-06  
 ; PRIOR APPLICATION NUMBER: 60/279,526  
 ; PRIOR FILING DATE: 2001-03-27  
 ; NUMBER OF SEQ ID NOS: 8481  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO: 298  
 ; LENGTH: 1085  
 ; TYPE: DNA  
 ; ORGANISM: Bacillus licheniformis  
 ; LENGTH: 1085  
 ; US-09-974-300-298

RESULT 5  
US-10-083-357-579 ; Sequence 579, Application US/10083357  
; Publication No. US20030054370A1  
; GENERAL INFORMATION  
; APPLICANT: Olandong Zeng et al.  
; TITLE OF INVENTION: Systemic Discovery of New Genes  
; FILE REFERENCE: 03796-090  
; CURRENT APPLICATION NUMBER: US/10/083,357  
; CURRENT FILING DATE: 2002-02-27  
; NUMBER OF SEQ ID NOS: 1346  
; SEQ ID NO 579  
; LENGTH: 198  
; TYPE: DNA  
; ORGANISM: *Saccharomyces cerevisiae*  
; US-10-083-357-579

Query Match 3.5%; Score 37.8; DB 9; Length 198;  
Best Local Similarity 57.0%; Pred. No. 0.11; Gaps 0;  
Matches 69; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 724 TCTCAGGTGATGCCAAAAGTTCCTCACGGAGAGAAATTCTTAATGGAAAAGATA 783  
DB 46 TTAAAGCGGGGAGGCCAAAAGTCTCTGAAAGCAATGCAAGCTAACATGCTTAAGATA 105  
QY 784 CAGGAGGTTGTAAACAGCGAGAAAAGGATGTAGCAGCAAATGAAAGGAGTT 843  
DB 106 TGGTTGGTTAGAAGTGCGTCAATTATGCATCTCGATGAGCAAATATACGATCT 165

QY 844 G 844  
DB 166 G 166

Query Match 3.5%; Score 37.6; DB 10; Length 588;  
Best Local Similarity 50.6%; Pred. No. 0.24; Gaps 0;  
Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 680 CAGAAATTACAAGGTGAGACTCCCTCAACTGACCTACCCCTCTAACGTGACCTACGGTGTGGCAA 739  
Db 309 CAGGCCCTTAACTGGCAAAATGTGCCAACTATGGCAAAATGTGGTAACTTGTGGTACTTA 368  
QY 740 AAGTCTCCCTCACGGAGGAAATCTAATGGAAAAGATAACAGGTTGTAACAG 799  
Db 369 CAGAAGAAGAAGTGGGGAGAAAGGATATTGAAAACAAAAATAATGGCTGAAACT 428  
QY 800 AGCGAGAAAGGATCTACGCAAATGAAAGAAAATAGAGGTTGAGTTTCGGAGTA 859  
Db 429 TCCCCAAATCTGATGAAAGCATATAATACATCCAAAGTATTGGAAGTCCAAAGTA 488

RESULT 7  
US-09-904-761-6646 ; Sequence 6646, Application US/09864761  
; Patent No. US20020054376A1  
; GENERAL INFORMATION  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Acomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-16  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263,6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359

Sequence 365, Application US/08781986A  
; Publication No. US2003005436A1  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: *Staphylococcus aureus* Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue

CITY: Rockville  
 STATE: Maryland  
 COUNTRY: USA  
 ZIP: 20850

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Disette, 3.50 inch, 1.4Mb storage  
 COMPUTER: HP Vectra 486/33  
 OPERATING SYSTEM: MS DOS version 6.2  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/781,986A  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Benson, Bob  
 REGISTRATION NUMBER: 30,446  
 REFERENCE/DOCKET NUMBER: PB248PP  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (301) 309-8504  
 TELEFAX: (301) 309-8512  
 INFORMATION FOR SEQ ID NO: 365:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 6804 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear

US-08-781-986A-365

Query Match Score 3.4%; Best Local Similarity 57.5%; Pred. No. 2.7; DB 7; Length 6804; Indels 0; Gaps 0; Matches 65; Conservative 0; Mismatches 48; Del 0; Gaps 0;

Qy 337 GCGAAAGTCTCGGAAGAACCCCTCGTAAGGTCAACTTGACCAAACTTGATATGAC 396  
 Db 644 GCTCAGATTATAAAGTCGCCATGAAACTGATTAATGACCATATTGATTAGAT 703

Qy 397 TTGAAAGAAGTATGAAATTAGAGAAAACCGTTCGGTACFT 449  
 Db 704 GAAGGAAACGTACTGAAATGAAAAATGAAATCTATGAGGGTCAT 756

RESULT 8  
 US-09-974-300-4716  
 Sequence 4716, Application US/09974300  
 Patent No. US2002146721A1  
 GENERAL INFORMATION:  
 APPLICANT: Berka, Randy M.  
 TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
 TITLE OF INVENTION: Expression  
 FILE REFERENCE: 10085, 500 US  
 CURRENT APPLICATION NUMBER: US/09/974,300  
 CURRENT FILING DATE: 2001-10-05  
 PRIOR APPLICATION NUMBER: 09/680,598  
 PRIOR FILING DATE: 2000-10-06  
 PRIOR APPLICATION NUMBER: 60/279,526  
 NUMBER OF SEQ ID NOS: 8481  
 SOFTWARE: FastSEQ for Windows Version 4.0  
 SEQ ID NO 4716  
 LENGTH: 560  
 TYPE: DNA  
 ORGANISM: Bacillus clausii  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (1)...(560)  
 OTHER INFORMATION: n = A,T,C or G

US-09-974-300-4716

Query Match Score 3.4%; Best Local Similarity 45.1%; Pred. No. 0.84; Matches 133; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

Qy 154 TACCCAGACCCCCAAGCAGAAAGTAAAGGGTCTTGGCGATTTTGGCGTAAG 213  
 Db 181 TAUCGGATGGCTATGGCACGCTTGCCAAAGCATAAAGAGAT 240

Qy 214 GAAGAAAATTAGTTAGTCTGGTAACGGTICGGACAACTCATATACCTCTCAATAGCT 273  
 Db 241 GAAGGACAGCTTATAGGAAATGGACCGATGAATTCACACTAGCAGAAGT 300

Qy 274 ATAGGTGAACTTACACATACCGTTAGTACACCTGTCACCTTCATGAGAGATA 333  
 Db 301 TTTTAAGGCCTGAGACCAACAGGTTATGGGACGCCCTTCTCAGTACAAGTTA 360

Qy 334 AGTGCGAAAGTCTCGGAAGAACCCCTCGTAAGGTAACTGGACAAAACCTTGATATA 393  
 Db 361 AATGCAACGATGGGCAACGATTAAGAAATAACCGTTAAAGAACGGCTCCAC 420

Qy 394 GACTTAGAAAGAAGTATTGAAATTAGAGAAAACCGTTCTGGTACT 448  
 Db 421 GATTAGATGCAATGGCTTGACGCCAAACGCGATTGACRCAAAACGCGATTGCTGGCT 475

RESULT 9  
 US-09-738-626-2298  
 Sequence 2298, Application US/09738626  
 Publication No. US2002197605A1  
 GENERAL INFORMATION:  
 APPLICANT: NAKAGAWA, SATOSHI  
 APPLICANT: MIZOGUCHI, HIROSHI  
 APPLICANT: ANDO, SEIKO  
 APPLICANT: HAYASHI, MIKIRO  
 APPLICANT: OCHIAI, KEIKO  
 APPLICANT: YOKOI, HARUHIKO  
 APPLICANT: TATEISHI, NAOKO  
 APPLICANT: SENOH, AKIHIRO  
 APPLICANT: IKEDA, MASATO  
 APPLICANT: OZAKI, AKIO  
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 CURRENT APPLICATION NUMBER: US/09/738,626  
 PRIOR APPLICATION NUMBER: JP 99/377484  
 PRIOR FILING DATE: 2000-12-18  
 FILE REFERENCE: 249-12-15  
 CURRENT FILING DATE: 2000-04-07  
 PRIOR APPLICATION NUMBER: JP 00/159162  
 PRIOR FILING DATE: 2000-04-07  
 PRIOR APPLICATION NUMBER: JP 00/280988  
 NUMBER OF SEQ ID NOS: 7059  
 SEQ ID NO 2298  
 LENGTH: 1098  
 TYPE: DNA  
 ORGANISM: Corynebacterium glutamicum  
 US-09-738-626-2298

Query Match Score 3.4%; Best Local Similarity 45.5%; Pred. No. 1.2; Matches 127; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

Qy 235 AACGGTTCGGCGAACCTACATACCTCTCAATAGCTTACATACCC 294  
 Db 286 AATGGTCCCATGAAATTCTCGCAGGCCTTGGCTGACCTGGCAGC 345

Qy 295 GTTACATACCTGTCGCCACCTTCCATGAGATAAGTGGAAACTTCCTGGAGA 354  
 Db 346 GCGTGGGATTCACGCCGCTATTCATCACCCATTGGCTAAAGGCCACCACT 405

Qy 355 CCCCTGGTAAGGTCACCTGGACAAAAACTTGTACATGACTAGAAAGAATSTATGAA 414  
 Db 406 GAATTCAATTGGGTCCCAGGTCTGATTCCGCATCCTGATGTCGGCTGGAA 465

Qy 415 TTAATAGAGAAAGAAAACCCGGTCTCGGGTACTTGTACCCAAACACCCACGGGA 474  
 Db 466 GAAATCCTGTTGAAAGAGCCGTCACATGTTGTCACCCGAAACCCGACCCGT 525

Qy 475 ACCCTCTTCCAGGGAAAGATGGAGGATAAGAAC 513  
 Db 526 GATGTGACCTCGTGACGATGTTAGCGCATCATCAC 564

---

**RESULT 10**  
 US-09-971-536-8  
 ; Sequence 8, Application US/09971536  
 ; Patent No. US20020159976A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Glenn, Matthew  
 ; Haukala, Tirkka  
 ; Bloksberg, Leonard  
 ; Lubbers, Mark  
 ; Dekker, James  
 ; Christansson, Anna  
 ; Holland, Ross  
 ; O'Toole, Paul  
 ; Reid, Julian  
 ; APPLICANT: Coolbear, Timothy  
 ; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods  
 ; TITLE OF INVENTION: Using Them  
 ; FILE REFERENCE: 1043c2  
 ; CURRENT APPLICATION NUMBER: US/09/971-536  
 ; CURRENT FILING DATE: 2001-10-02  
 ; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238  
 ; PRIOR FILING DATE: 2000-08-08  
 ; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623  
 ; PRIOR FILING DATE: 2000-11-28  
 ; PRIOR APPLICATION NUMBER: PCT/NZ01/00160  
 ; PRIOR FILING DATE: 2001-08-08  
 ; NUMBER OF SEQ ID NOS: 83  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 8  
 ; LENGTH: 1350  
 ; TYPE: DNA  
 ; ORGANISM: Lactobacillus rhamnosus  
 US-09-971-536-8

Query Match 3.38; Score 35.2; DB 9; Length 1350;  
 Best Local Similarity 46.1%; Pred. No. 2.1; Indels 3; Gaps 1;  
 Matches 155; Conservative

Qy 152 ATAPACCAGACCCGAGCGAAAGAGTTAAAGCGGTTCCTGCGATTTCGGCTTA 211  
 Db 304 ATTACCCGAGATGGCTATGCACTCAACTGGACCCGGTGGCAAGGTAGCGTTG 363

Qy 212 AGGAGAAATTAATGAGCTGGAGACTATACTCTCAATAG 271  
 Db 364 CGCGAGGAGCTGGTTTGGCAAGGTGGCTGAGTAAGCTTGCCTTAATGCCGCA 423

Qy 272 CTATAGGTGAACTTACATCCGGTTACATACCTGTCACCTTCCATGTACGAGA 331  
 Db 424 CTTTTAGCCGGGGATGAAAGGGCACCTGATGACATGACCGT 483

Qy 332 TRAGTGCAGAAGTCTCGGAAGACCCCTCGTAAAGGTCAACTGGACAAAATTTGATA 391  
 Db 484 TGCGATGCCAAATGAAAGGGCACCTGATGATGTCGCCCTCACTGAAACTGGCAATT 543

Qy 392 TAGACTTAGAAGAAGTATGATAATAGRAAAGAAAACCGPCTCGGTATTTG 451  
 Db 544 TTGTATTATCTGAATGCCAGGGCTCACTGATGTCGCCCTCACTGAAACTGGGTGT 600

Qy 452 CTTACCCAAAACACCCACGGAAACCTCTTCCAA 487  
 Db 601 GCAACCAAATAACCCACGGACGGTGTGCAA 636

US-09-938-842A-3936/c  
 ; Sequence 3936, Application US/09938842A  
 ; Patent No. US2002010378A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Harper, Jeff  
 ; Keeps, Joel  
 ; APPLICANT: Wang, Xun  
 ; APPLICANT: Zhu, Tong  
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
 ; TITLE OF INVENTION: SAME, AND METHODS OF USE  
 ; FILE REFERENCE: SCRIP1300-3  
 ; CURRENT APPLICATION NUMBER: US/09/938-842A  
 ; CURRENT FILING DATE: 2001-08-24  
 ; PRIORITY NUMBER: US 60/227,866  
 ; PRIORITY FILING DATE: 2000-08-24  
 ; PRIORITY APPLICATION NUMBER: US 60/264,647  
 ; PRIORITY FILING DATE: 2001-11-16  
 ; PRIORITY APPLICATION NUMBER: US 60/300,111  
 ; PRIORITY FILING DATE: 2001-06-22  
 ; NUMBER OF SEQ ID NOS: 5379  
 ; SEQ ID NO 3936  
 ; LENGTH: 2000  
 ; TYPE: DNA  
 ; ORGANISM: Arabidopsis thaliana  
 US-09-938-842A-3936

Query Match 3.28; Score 34.6; DB 9; Length 2000;  
 Best Local Similarity 59.8%; Pred. No. 4.1; Indels 0; Gaps 0;  
 Matches 58; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 822 AATGAGAAATAGAAGGAGTTGAGGTTCCTGGATTAAGGTTACTCTTGCCTTTCAG 881  
 Db 1625 AAAACAGAGTTGAGAAGTTGCAAGTCTCCAAATAAAGGTCAACTCTGTATTTCAC 1566

Qy 882 AACGCCTTACCCGCCACGAGTTATCAGGAGCTA 918  
 Db 1565 AAGTTTAAATCIGCTAAATAGGTTATGATCATCTA 1529

**RESULT 12**  
 US-09-864-761-11114  
 ; Sequence 11114, Application US/09864761  
 ; Patent No. US2002004876A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharron G.  
 ; Hanzel, David R.  
 ; APPLICANT: Rankin, David R.  
 ; APPLICANT: Chen, Weisheng  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
 ; FILE REFERENCE: Aeomica-X-1  
 ; CURRENT APPLICATION NUMBER: US/09/864,761  
 ; CURRENT FILING DATE: 2001-05-23  
 ; PRIOR APPLICATION NUMBER: US 60/180,312  
 ; PRIOR FILING DATE: 2000-02-04  
 ; PRIOR APPLICATION NUMBER: US 60/207,456  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: US 09/632,366  
 ; PRIOR FILING DATE: 2000-08-03  
 ; PRIOR APPLICATION NUMBER: GB 24263-6  
 ; PRIOR FILING DATE: 2000-10-04  
 ; PRIOR APPLICATION NUMBER: US 60/236,359  
 ; PRIOR FILING DATE: 2000-09-27  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665  
 ; PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668 ; TYPE: DNA  
; PRIOR FILING DATE: 2001-01-30 ; ORGANISM: Arabidopsis thaliana  
; PRIOR APPLICATION NUMBER: PCT/US01/00663 ; US-09-887-576-455  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662 ; Query Match 3.2%; Score 34.4; DB 10; Length 2270;  
; PRIOR FILING DATE: 2001-01-30 Best Local Similarity 54.8%; Pred. No. 5,1;  
; PRIOR APPLICATION NUMBER: PCT/US01/00661 Matches 68; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1  
; SEQ ID NO: 11114  
; LENGTH: 375  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC006212.1  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.65  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.49  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.58  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.77

US-09-864-761-11114

Query Match 3.2%; Score 34.4; DB 10; Length 375;  
Best Local Similarity 53.8%; Pred. No. 1,8;  
Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 921 GAAAGGGATGTCCTCGAGAACGTTACATGGAAAGGACTCAAAGTGCCTCAG 980  
Db 67 GAGAGAGAGACGGTCCTTCATTCTGAAATGGAACTGACAG 126  
Qy 981 GGTAAAGCGTAGGAAACCGAGAAAACAACAAAGTTCTGAAGCAGTGAGGAGAT 1040  
Db 127 TGTGAATCTGCCAACACAGAAACCATACAGTGCTGAGTAAAGAATAGGA 186  
Qy 1041 AAAATCCCTTC 1052  
Db 187 AAGGACCCCTGTC 198

RESULT 14  
US-09-927-668-11

Sequence 11, Application US/09927668  
; Patent No. US2002011514A1  
; GENERAL INFORMATION:  
; APPLICANT: Kaplan, Jerry  
; PEROU, Charles  
; MOORE, Karen  
; TITLE OF INVENTION: COMPOSITIONS FOR THE DIAGNOSIS  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 115 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036/2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSSO Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/927,668  
; FILING DATE: 10-Aug-2001  
; CLASSIFICATION: <Unknown>  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 09/396,540  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-062-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-869-9090  
; TELEFAX: 212-869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12225 base pairs  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 190..11208  
; ; SEQUENCE DESCRIPTION: SEQ ID NO: 11:  
; US-09-927-668-11

Query Match 3.2%; Score 34.4; DB 10; Length 12225;

RESULT 13  
US-09-887-576-455/c  
; Sequence 455, Application US/09887576  
; PATENT NO. US20020144047A1  
; GENERAL INFORMATION:  
; APPLICANT: Budworth, P.  
; APPLICANT: Brown, D.  
; APPLICANT: Chang, H.  
; APPLICANT: Zhu, T.  
; APPLICANT: Han, B.  
; APPLICANT: Wang, X.  
; APPLICANT: Cooper, Bret  
; TITLE OF INVENTION: Promoters for regulation of plant expression  
; FILE REFERENCE: 1360\_001US1  
; CURRENT APPLICATION NUMBER: US/09/887,576  
; CURRENT FILING DATE: 2001-06-25  
; PRIOR APPLICATION NUMBER: US 60/213,848  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: US 60/258,692  
; PRIOR FILING DATE: 2000-12-29  
; NUMBER OF SEQ ID NOS: 875  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO: 455  
; LENGTH: 2270

Best Local Similarity	Pred.	No.	13;	Indels	0;	Gaps	0;	Matches	98;	Conservative	0;	Mismatches	106;	Indels	0;	Gaps	0;
Matches	98;	Conservative	0;	Mismatches	106;	Indels	0;	Matches	98;	Conservative	0;	Mismatches	106;	Indels	0;	Gaps	0;
Qy	102	GGAGATAAAACAGGGCCTTAGAAAATTAAAAAGGTCCCCCTGAAACAATACCCAGA	161					Qy	102	GGAGATAAAACAGGGCCTTAGAAAATTAAAAAGGTCCCCCTGAAACAATACCCAGA	161						
Db	8430	GGATICAATCAAGGATATATCTAAATACCTGCAAGATATACCCAGGCACTGTCTCT	8489					Db	8817	GGATICAATCAAGGATATATCTAAATACCTGCAAGATATACCCAGGCACTGTCTCT	8876						
Qy	162	CCCGAGGCAAAGACTTAAGGCCCTTCTGGGAATTTCGGCTTAAGGAGAAAA	221					Qy	162	CCCGAGGCAAAGACTTAAGGCCCTTCTGGGAATTTCGGCTTAAGGAGAAAA	221						
Db	8490	CTCCCAAGGAAATGAGAAAAAAAGGTGATCCAGCATATAGGAATGATAAGTGA	8549					Db	8877	CUCCCAAGGAAATGAGAAAAAAAGGTGATCCAGCATATAGGAATGATAAGTGA	8936						
Qy	222	TTAGTCTCGGTAAAGGTGACACTATACCTCAATAGCTATAGTGA	281					Qy	222	TTAGTCTCGGTAAAGGTGACACTATACCTCAATAGCTATAGTGA	281						
Db	8550	TTCGAGTGCAGCAGACATTCAGGAACATTACAGACTGACATGATAAGCAGT	8609					Db	8931	TTGAGTGCAGCAGACATTCAGGAACATTACAGACTGACATGATAAGCAGT	8996						
Qy	282	ACTTTACATACCGTTACATACC	305					Qy	282	ACTTTACATACCGTTACATACC	305						
Db	8610	ATGGATGACCCATCTACTATCC	8633					Db	8997	ATGGATGACCCATCTACTATCC	9020						

RESULT 15  
US-09-927-668-9  
; Sequence 9, Application US/09927668  
; Patent No. US200115141A1  
; GENERAL INFORMATION:  
; APPLICANT: Kaplan, Jerry  
; ;  
; Moore, Karen  
; ;  
; TITLE OF INVENTION: COMPOSITIONS FOR THE DIAGNOSIS  
; AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036/2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/927,668  
FILING DATE: 10-Aug-2001  
CLASSIFICATION: <Unknown>  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/09/927,668  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-062-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12616 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 190..11592  
SEQUENCE DESCRIPTION: SEQ ID NO: 9:  
US-09-927-668-9  
Query Match 3.2%; Score 34.4%; DB 10; Length 12616;  
Best Local Similarity 48.0%; Pred. No. 14;

result	No.	Score	Query	Match	Length	DB	ID	Description
1	65	8	6.2	573	10	BE432349	EST398878	BE432349 EST398878
2	65	6.1	587	10	AW034522	EST77838	AW034522 EST77838	
3	62	6	819	12	BG350830	098G11	BG350830 098G11	
4	58	8	762	13	BM412207	EST586334	BM412207 EST586334	
5	58	5.5	631	13	BJ463860	BU463860	BJ463860 BU463860	
6	57	5.4	778	12	BG887294	EST513145	BG887294 EST513145	



AUTHORS	Nielsen,K.L., Crookshanks,M., Emmersen,J. and Welinder,K.G.
TITLE	EST sequencing of mature potato tuber (Var. Kuras)
JOURNAL	Unpublished (2000)
COMMENT	Contact: Karen G. Welinder Institut for bioteknologi Aalborg Universitet Sohngardsholmsvej 49, 9000 Aalborg, Denmark Tel: +45 96358467 Fax: +45 98141808 Email: kgw@bio.auc.dk Sequenced from the 5' end. High quality sequence stop: 819
FEATURES	<p>source</p> <p>1. .819 /organism="Solanum tuberosum" /cultivar="Field grown Kuras" /db_xref="Field taxon:4113" /clone_lib="Mature tuber lambda ZAP" /tissue_type="Tuber" /note="Vector: Lambda ZAP"</p> <p>BASE COUNT 227 a 155 c 194 g 243 t</p> <p>ORIGIN</p>
Query Match	5.9%; Score 62.6; DB 12; Length 819;
Best Local Similarity	47.0%; Pred. No. 2.6e-06;
Matches 194; Conservative	0; Mismatches 219; Indels 0; Gaps 0;
Qy	149 ACAAAATACCCAGAACCGGAAAGCGTTAAAGGGGTTCTGGGATTTCGGC 208
Db	393 ATATTTATCCGATCCGTAAAGCCGTACCGTGCACTCTGGAGATCTGGC 452
Qy	209 TAAAGGAAAGAAAATTAGTCTCGGTAAACGGTTCGACAACTATACCTCTCA 268
Db	453 TTGAATCTGGAGTATTCATGAGGGTGGGTGAGATGAACCTATTGATAATGA 512
Qy	269 TAGCTATAGGTGAACTTTACACCGTTACATACCTGTTCCACCATGTA 328
Db	513 GATGCCATTGATCTGGACAGAGATCTGGCACCCACTTCACAACTGTTG 572
Qy	329 AGATAAGTGCAGAAAGTCTCGGAAGAACCCCTCGTAAGGGTCACTGGAAACTTGT 388
Db	573 AATTGATGGAGCTGAAATGGAGCTAGGTCATCAAGGTGCCAGACTTA 632
Qy	389 ATATAGACTTAGAAGAAGTATTGATTATAGAAAGAAACCCGTTCTCGGTACT 448
Db	633 GCCTGGATGAGAAACGGATGCGGAAGTGGTGAACCTGAGATGCAATATTIC 692
Qy	449 TTGGCTTACCCAACACCCACGGGAAACCTCTTCACGGGAAAGATTGAGGAGATA 508
Db	693 TGACATCACCTAAATACCTGATGGAGTAAATGGATGAAACTTATGAAATAC 752
Qy	509 GAAACAGGGTGTCTGTGTAATAGACGAAGGCTACTATCATACTCGGA 561
Db	753 TTACACCTGCCCATATTGGATAATGGATGAAGGATATGACTTTCTGGA 805
RESULT 4	
LOCUS	BMA12207
DEFINITION	EST586534 tomato breaker fruit Lycopersicon esculentum cDNA clone CLEG59E12 5' end, mRNA sequence.
ACCESSION	BMA12207
KEYWORDS	EST, BMA12207, GI:18263837
SOURCE	tomato
ORGANISM	Lycopersicon esculentum Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon
REFERENCE	1 (bases 1 to 765)
AUTHORS	Alcalá,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai

J., Bougrí,O., Kirkness,E., Utterback,T., Van Aken,S., Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.	
Generation of ESTs from tomato fruit tissue, breaker stage (2002)	
Unpublished (2002)	
Contact: CIGI	
Clemson University Genomics Institute	
Clemson University	
100 Jordan Hall, Clemson, SC 29634, USA	
Email: <a href="http://www.genome.clemson.edu/orders/">http://www.genome.clemson.edu/orders/</a>	
This clone is available through the Clemson University Genomics Institute	
Seq primer: T3.	
FEATURES	Location/Qualifiers
source	1. .762 /organism="Lycopersicon esculentum" /note="Vector: phleuscript:Skmcuadapt; Site_1: EcorI;" /cultivar="TA496" /db_xref="Taxon:081" /clone="CLEG59E12" /clone_lib="tomato breaker fruit" /tissue_type="Pericarp" /dev_stage="breaker" /lab_host="SOLR"
site	Site 2: Xhol; supplier: Boyce Thompson Institute for Genomic Research. Fruit sequencing: The Institute for Genomic Research. Fruit were harvested at the breaker stage (first sign of lycopene accumulation on the blossom end of fruit). Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp.
BASE COUNT	207 a 155 c 178 g 222 t
ORIGIN	Query Match 5.5%; Score 58.8%; Pred. No. 2.9e-05; Best Local Similarity 48.8%; Mismatches 167; Indels 0; Gaps 0;
Qy	149 ACAAAATACCCAGAACCGGAAAGCGTTAAAGGGGATTTTCGGGATTTTCGGCG 208
Db	395 ACATTTATCCGTCAAAGCCGACGGCTGGTGAAGCTGATTCGATGCG 454
Qy	209 TTAAGGAAAGAAAATTAGTTCTCGGAACGGTCACTGGATTTCTCCATGTCAA 268
Db	455 TTGAATCTGGATATCTGAGTATCTGAGGGTGGTGCAGATGACTCATGATTGATAATGA 514
Qy	269 TAGCTATAGGTGAACTTTACACCGTTACATACCTGTTCCACCATGTA 328
Db	515 GATGCCATTGATCTGGTCAAGGATGTTGACTTCCCACTTCACAATGATG 574
Qy	329 AGATAAGTGCAGAAAGTCTCGGAAGAACCCCTCGTAAGGTCAACTGGACGAAACTTGT 388
Db	575 AAATTGATGGAGCTTAATGGACGGTCAAGGTGCCAGGTGCTAGTAACTGGCTAGAAACCGACTTA 634
Qy	389 ATATAGACTTAGAAGAAGTATTGATTATAGAAAGAAACCCGTTCTCGGTACT 448
Db	635 GCCTGGAGCTGAAACGGTAAACGGTAAAGGTGCTAGTAACTGGCTAGAAACCGACTTA 694
Qy	449 TTATGACTTAAAGGAGTATTGATTATAGAAAGAAACCCGTTCTCGGTACT 448
Db	695 GACATCACCTAACACCCACGGGA 474
RESULT 5	
LOCUS	BM412207
DEFINITION	EST586534 tomato breaker fruit Lycopersicon esculentum cDNA clone CLEG59E12 5' end, mRNA sequence.
ACCESSION	BMA12207
KEYWORDS	EST, BMA12207, GI:18263837
SOURCE	tomato
ORGANISM	Lycopersicon esculentum Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon
REFERENCE	1 (bases 1 to 765)
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon
J., Bougrí,O., Kirkness,E., Utterback,T., Van Aken,S., Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.	
Generation of ESTs from tomato fruit tissue, breaker stage (2002)	
Unpublished (2002)	
Contact: CIGI	
Clemson University Genomics Institute	
Clemson University	
100 Jordan Hall, Clemson, SC 29634, USA	
Email: <a href="http://www.genome.clemson.edu/orders/">http://www.genome.clemson.edu/orders/</a>	
This clone is available through the Clemson University Genomics Institute	
Seq primer: T3.	
FEATURES	Location/Qualifiers
source	1. .762 /organism="Lycopersicon esculentum" /note="Vector: phleuscript:Skmcuadapt; Site_1: EcorI;" /cultivar="TA496" /db_xref="Taxon:081" /clone="CLEG59E12" /clone_lib="tomato breaker fruit" /tissue_type="Pericarp" /dev_stage="breaker" /lab_host="SOLR"
site	Site 2: Xhol; supplier: Boyce Thompson Institute for Genomic Research. Fruit sequencing: The Institute for Genomic Research. Fruit were harvested at the breaker stage (first sign of lycopene accumulation on the blossom end of fruit). Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp.
BASE COUNT	207 a 155 c 178 g 222 t
ORIGIN	Query Match 5.5%; Score 58.8%; Pred. No. 2.9e-05; Best Local Similarity 48.8%; Mismatches 167; Indels 0; Gaps 0;
Qy	149 ACAAAATACCCAGAACCGGAAAGCGTTAAAGGGGATTTTCGGGATTTTCGGCG 208
Db	395 ACATTTATCCGTCAAAGCCGACGGCTGGTGAAGCTGATTCGATGCG 454
Qy	209 TTAAGGAAAGAAAATTAGTTCTCGGAACGGTCACTGGATTTCTCCATGTCAA 268
Db	455 TTGAATCTGGATATCTGAGTATCTGAGGGTGGTGCAGATGACTCATGATTGATAATGA 514
Qy	269 TAGCTATAGGTGAACTTTACACCGTTACATACCTGTTCCACCATGTA 328
Db	515 GATGCCATTGATCTGGTCAAGGATGTTGACTTCCCACTTCACAATGATG 574
Qy	329 AGATAAGTGCAGAAAGTCTCGGAAGAACCCCTCGTAAGGTCAACTGGACGAAACTTGT 388
Db	575 AAATTGATGGAGCTTAATGGACGGTCAAGGTGCCAGGTGCTAGTAACTGGCTAGAAACCGACTTA 634
Qy	389 ATATAGACTTAGAAGAAGTATTGATTATAGAAAGAAACCCGTTCTCGGTACT 448
Db	635 GCCTGGAGCTGAAACGGTAAACGGTAAAGGTGCTAGTAACTGGCTAGAAACCGACTTA 694
Qy	449 TTCTCTACCCAAACACCCACGGGA 474
Db	695 GACATCACCTAACACCCACGGGA 720
RESULT 5	
LOCUS	BJ463860
DEFINITION	mRNA linear EST 23-MAY-2002
ACCESSION	BJ463860
KEYWORDS	EST, BJ463860 K. Sato unpublished cDNA library, cv. Haruna Nijo germination shoots Hordeum vulgare subsp. vulgare cdNA clone bags30m21 5', mRNA sequence.
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	

REFERENCE	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae 1 (bases 1 to 631)	AUTHORS	van der Hoeven, R.; Bezzelides, J.; Ewing, E.; Cho, J.; Chiemingo, A.; Bougrin, O.; Bueli, C.R.; Ronning, C.; Tantsky, S. and Baker, B.
AUTHORS	Sato, K., Saisho, D. and Takeda, K.	TITLE	Generations of ESTs from dormant potato tubers
JOURNAL	Barley EST sequencing project in NIG and Okayama Univ Unpublished (2002)	JOURNAL	Unpublished (2001)
COMMENT	Contact: Tadao Shin-i Center For Genetic Resource Information National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6855 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp.	COMMENT	Contact: Cathy Ronning The Institute for Genomic Research For clone info: Please contact Research Genetics, Libraries Division tel 1-800-711-6195, email cdna@resgen.com Seq primer: M13F-R.
FEATURES	source	Location/Qualifiers	Location/Qualifiers
FEATURES	source	source	/organism="Solanum tuberosum" /cultivar="Kennebec" /db_xref="Taxon:4113" /clone="CSTD4013"
FEATURES	source	/clone.lib="CSTD" /tissue_type="dormant tuber" /dev_stage="one month post-harvest"	/clone.lib="CSTD" /tissue_type="dormant tuber" /dev_stage="one month post-harvest"
FEATURES	source	/lab.host="SOLR"	/lab.host="SOLR"
FEATURES	source	/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XbaI; This library targets genes expressed in dormant tubers. This library was made from sections of dormant tuber, avoiding the buds and epidermis. Tubers were stored for one month post-harvest at 4°C. The tuber was peeled, well away from the surface. Then it was chopped into 1-2 mm cubes and immediately frozen in liquid nitrogen. This library is noted as P4 in Tanksley lab notebooks."	/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XbaI; This library targets genes expressed in dormant tubers. This library was made from sections of dormant tuber, avoiding the buds and epidermis. Tubers were stored for one month post-harvest at 4°C. The tuber was peeled, well away from the surface. Then it was chopped into 1-2 mm cubes and immediately frozen in liquid nitrogen. This library is noted as P4 in Tanksley lab notebooks."
BASE COUNT	172 a 128 c 145 g 185 t 1 others	BASE COUNT	214 a 153 c 181 g 230 t
ORIGIN		ORIGIN	
Query Match	5.5%; Score 58.4; DB 13; Length 631;	Query Match	5.4%; Score 57.2; DB 12; Length 778;
Best Local Similarity	46.5%; Pred. No. 3.5e-05;	Best Local Similarity	48.5%; Pred. No. 8e-05;
Matches	188; Conservative 0; Mismatches 216; Indels 0; Gaps 0;	Matches	158; Conservative 0; Mismatches 168; Indels 0; Gaps 0;
QY	154 TACCCAGACCCCCGAAAGCGTTAAAAGAGTTTCGGGATTTCGGCTTAAG 213	QY	149 ACAATAACCAGACCCCCGAAAGCGTTAAAAGAGTTTCGGGATTTCGGATTTTCGGG 208
Db	85 TACCCATGATCCGAAAGCCGCACGTGGCTGCGTGTAGCTTAATGAGATGT 204	Db	399 ATATTATCTGATCTGAGCTGAGCTGCTTGTCAGAAGTTCGGCC 458
QY	214 GAAGAAATTAGTGTCTCGGTAACGGTTCGGAAGAACCTCATACTACCTCTCAATAGCT 273	QY	209 TTANGGAAGAAAATTACTCTCGGTAAGGGTTCGAGAACCTCATACCTACCTCTCAA 268
Db	145 ACTGAGCACATCTGTGGATGGCTGCGTGTAGCTTAATGAGATGT 204	Db	459 TTGATCTCTGATATTCGGAGGTGGTCAAGTAACTGATGATGATATGTA 518
QY	274 ATAGTAGAACCTTACATACCCCTTACATACCTGTCCACCTTCCCATGAGAGATA 333	QY	269 TAGCTATAGTGAACTTACATACCCGTTACATACCTGTCCACCTTCCCAC 328
Db	205 GTCATCTGAGCTGATAATTGTGATGCTTACACTTACCATGTAACATT 264	Db	519 GATCGCTATGGATCCCTGGTACAGATGTTGACTGCCACCCACTTCACAACTGTATG 578
QY	334 ATGCGAAAGTCTCGGAAAGCCCTCGTAAGGGTCAACTGTGAGAAACCTGTATA 393	QY	329 AGATAAGTGGAAAGTCTCGGAAAGCCCTCGTAAGGGTCAACTGTGAGAAACCTTG 388
Db	265 GATGCTCTGAGCTAATGTGCACTTGTGICATCAAAGTCCCANAGACTCTGTATTCTCCCATA 324	Db	579 AATITGATCTGAGCTGATTAATGGCAGCTGTCATCAAGTGGCTAGGACACTTA 638
QY	394 GACTTGTAAAGAAGTATTGAAATTGAGAAAGAAAAACCGTCTCGGACTCTGTG 453	QY	389 ATAGACTTAGAAAAGTGAATTGAAATGAGATGAGATGAGATAAAACCCCTCTGGTACT 448
Db	325 GACATTGCAAAATTGTGCAATTGTTGAAGGGCAAAATGCAATTGACA 384	Db	639 GCCUGGATGATGAGGGATGCTGGCAAGTGTGAACTTGAGAACCGAGTGCATATTC 698
QY	454 TACCCAAACCAACCCACGGGAACCTCTTTCAGGGAAAGATGAGGAGATAAGAAC 513	QY	449 TTGCTTACCCAAACACCCACGGGA 474
Db	385 TCTCCGGCAACCCAGATGGCTGAAATCACGATGAGCTTCAGCTGAC 444	Db	699 GACATCACCTAATACTCGATGGGA 724
QY	514 AGGGTGTCTGTGTAATAAGCAAGAACCTACTATCTC 557	RESULT	7
Db	445 CTTCCGGFACTGTGAGCTGGCTGAGCAAGCATGGTACTTC 488	LOCUS	BG598806
DEFINITION	EST513145 cDNA clone cSTD1013' sequence.	DEFINITION	EST503076 cDNA clone cSTD22c22 5' sequence.
ACCESSION	BG887294	ACCESSION	BG598806
VERSION	EST513145	VERSION	EST503076
KEYWORDS		KEYWORDS	
SOURCE	potato.	SOURCE	EST.
ORGANISM	Solanum tuberosum	ORGANISM	Solanum tuberosum
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		Eukaryota; Magnoliophyta; Embryophyta; eudicots; Solanaceae; Solanum.
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE	1 (bases 1 to 778)	REFERENCE	1 (bases 1 to 811)

AUTHORS	van der Hoeven,R., Bezzerezides,J., Sun,H., Cho,J., Chiemingo,A., Bougrini,O., Bueli,C.R., Ronning,C., Tanksley,S. and Baker,B.	TITLE	An EST database from Sorghum: plants infected with a compatible pathogen
JOURNAL	Unpublished (2000)	JOURNAL COMMENT	unpublished (2002) Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
COMMENT	Contact: Cathy Ronning The Institute for Genomic Research For clone info: Please contact Research Genetics, Libraries Division tel 1-800-711-6195, email cdna@resgen.com Seq primer: M3F-R.		
FEATURES source	Location/Qualifiers 1. 811 /organism="Solanum tuberosum" /cultivar="Kennebec" /db_xref="taxon:4113" /clone="csts2C22" /clone_idb="csts" /tissue_type="sprouting eyes from tubers" /dev_stage="12-14 weeks post harvest" /lab_host="SOLR" /note="Vector: pBluescript SK(-); Site_1: ECOR1; Site_2: XbaI; Various sizes of sprouting eyes (2mm to 15mm) were taken from tubers. The tubers were incubated at 26°C in the dark for 2-3 weeks prior to sprouting. The eyes were frozen in liquid nitrogen immediately upon removal from tubers." BASE COUNT 227 a ORIGIN 144 c 198 g 242 t	FEATURES source	TITLE
	Query Match Best Local Similarity 5.0% Matches 159; Conservative 0; Mismatches 176; Indels 0; Gaps 0;		JOURNAL
Qy	149 ACAAAATCCAGAACCGAGCGAAAGACGTTAAAGCGCTTCCTGCATTTCGGCG 208	Qy	5.0% Score 53.4; DB 12; Length 811;
Db	368 ATATTTATCCGTGAAAGCGGTAACTTGGCTGGCAACTCTGGTGAAGATTCTGGC 427	Db	478 ATGGAAAGAAATTAGTAGTCCTCGTAAGGTTGACACTCATACCTCTCAA
Qy	209 TAAAGGAAGAAATTAGTAGTCCTCGTAAGGTTGACACTCATACCTCTCAA 268	Db	428 TTGAAATCGAGTATTCTGCAGGGTGGGTGAGATGAACTCATGGTATGATA
Qy	269 TAGTATAGTGAACTTACATACCCGTTACATACCTGTTCCACCTTGTACG 328	Qy	489 GATGGTATGGATCTGGTGAAGATTGTTGACTGCCACCCACATTGATPAG 547
Db	488 GATGGTATGGATCTGGTGAAGATTGTTGACTGCCACCCACATTGATPAG 547	Qy	329 AGAAGTGGAAAGTCTCGGAAAGACCCCTCGTAAGGTTCAACTTGACGAAAACCTTG 388
Qy	548 AATTGATCGAGCTGTTAATGGGAGCTGTCAAGGTGCTAGAACCCAGACTTA 607	Db	548 AATTGATCGAGCTGTTAATGGGAGCTGTCAAGGTGCTAGAACCCAGACTTA 607
Qy	389 ATATAGACTTAGAAAGAATGAAATTGAAATGAGAAAACACCGTCTCGGGTACT 448	Qy	389 ATATAGACTTAGAAAGAATGAAATTGAAATGAGAAAACACCGTCTCGGGTACT 448
Db	608 GCCTGGATGAGAACGGATGGCAAGGGTGAACCTTGAACCGAAACTCTTT 483	Db	608 GCCTGGATGAGAACGGATGGCAAGGGTGAACCTTGAACCGAAACTCTTT 483
Qy	449 TTGGTTACCCAACACCCACGGAAACCTCTTT 483	Qy	449 TTGGTTACCCAACACCCACGGAAACCTCTTT 483
Db	668 TGACATACCTAAATCCTGTTGGGAGTAAATT 702	Db	668 TGACATACCTAAATCCTGTTGGGAGTAAATT 702
RESULT 8		RESULT 8	
LOCUS	BM322300	LOCUS	BM322300
DEFINITION	PC1_2-G06_b1-A002 mRNA linear EST 04-JAN-2002	DEFINITION	PC1_2-G06_b1-A002 Pathogen-infected compatible 1 (PIC1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION	BM322300	ACCESSION	BM322300
VERSION	GI:18059151	VERSION	GI:18059151
KEYWORDS	EST.	KEYWORDS	EST.
SOURCE	sorghum, bicolor	SOURCE	sorghum, bicolor
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum	ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum
REFERENCE	1 (bases 1 to 622)	REFERENCE	1 (bases 1 to 622)
AUTHORS	Cordonnier-Pratt,M.-M., Gingole,A., Fang,G.C., Dean,R., Wing,R., Sudman,M. and Pratt,L.H.	AUTHORS	Cordonnier-Pratt,M.-M., Gingole,A., Fang,G.C., Dean,R., Wing,R., Sudman,M. and Pratt,L.H.



Qy	334	AGTGGCAAAGTTCTCGGAAGACCCCTCGTAAGGTTCAACTGGACGAAACATTGTATAA	393	Query Match 4.78; Score 50.2; DB 13; Length 405;
Db	412	GACGTTCAGTCAAGTGTGACTTGTCAAGTCCAGACTPTCCATTCTCCCA	471	Best Local Similarity 47.5%; Pred. No. 0.0058; Matches 180; Conservative Matches 198; Indels 1; Gaps 1;
Qy	394	GAATTAGAAAAGTATTGAAATTAATAGAAAGAAAAACCGTTCTCGGTACTTTGC	453	Qy 183 AGCGGTCGTTGCGGATTTTCGCGTTAAGGAAGAAAATTAGTCTCGGAACCGTTC 242
Db	472	GACATGCAAACATGTCAAAATGGTGAACAGGAAACSCAAATGGCATATTGACA	531	Db 18 AGGGCACATTGCCAAAGATTCGGCTCAGAACNTCTGATTATTCATGCTGGTGTGTC 77
Qy	454	TACCCAAACAAACCCCACGGAA	475	Qy 243 GGAGCAACTCATACTACCTCTCAATAGCTATAGTGAAACTTACATACCGTTACAT 302
Db	532	TCTCCAAACAAACCCCTGATGCCA	553	Db 78 AGATGAACTATGTTGATTGGCATGTTGTTAGAACCTGGTAAACATTGGAA 137
<b>RESULT 11</b>				
BM173372	BM173372	405 bp mRNA linear EST 04-DEC-2001		Qy 303 ACCTGTTCCACCTTCCCATGTTACGAGATAAGTGGAAAAGTTCTGGAGAACCCCTCGT 362
LOCUS	DEFINITION	900/26 Avicennia marina leaf cDNA Library Avicennia marina cDNA clone Am900726 5' similar to histidinol phosphate aminotransferase (AC084320) of Nicotiana plumbaginifolia, mRNA sequence.		Db 138 CTCGCCACCAACATTAGAATTTAGTGGAGGTAACTGCTGCACTTTAT 197
ACCESSION	VERSION	BM173372	GI:17312935	Qy 363 AAAGGTCAACTGGCAAACACTTGTATAGACTTAGAAGAAAGTATTGATTANTAGA 422
KEYWORDS	EST			Db 198 TAAAGTACCTCGAAAGTCAGACTTCACCTGTTGATGTAAACGGATGTTGAGATGTTGA 257
SOURCE	ORGANISM	Avicennia marina.		Qy 423 GAAAGAAAACCGTTCTGGGTACTTTGCTTACCCAAACACCCACGGGAAACCTCTT 482
REFERENCE	Asteridae: euasterids I; Lamiales; Avicenniaceae; Avicennia.	Eukaryota; Viridiplantae; Streptophytta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae: euasterids I; Lamiales; Avicenniaceae; Avicennia.		Db 258 GCAGGAGAAACCAAAATGTTGATATTCTGACCTCTCACCTAAATACCCGATGGAAGCTTAAT 317
AUTHORS	Parani, M., Mehta, P., Sivapakash, K.R. and Parida, A.			Qy 483 TTCCAGGGAAAGATGGAGATAAGAACAGGGGTGTTCTGTAAATAGACCAAGC 542
TITLE	Expressed sequence tags from the mangrove species Avicennia marina			Db 318 CGATGAGAAACTCTTGGAAATTCCTAA-TCTGCCATATTGGTTGCTGGATAAAGC 376
JOURNAL	Unpublished (2000)			Qy 543 CTACTATCATTACTCCGGA 561
COMMENT	Contact: Parani, M / Parida, A			Db 377 ATACATGAAATTTCAGGA 395
<b>RESULT 12</b>				
BT17525	LOCUS	1031020D10.y1 C. reinhardtii CC-1690, Stress II (normalized), mRNA sequence.		Qy 511 mRNA 511 bp linear EST 19-SEP-2001
DEFINITION		Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.		Db BT17525 1031020D10.y1 C. reinhardtii CC-1690, Stress II (normalized), mRNA sequence.
ACCESSION				Db BT17525 1031020D10.y1 C. reinhardtii CC-1690, Stress II (normalized), mRNA sequence.
VERSION				Db BT17525 1 GI:15693220
KEYWORDS				Source EST.
SOURCE	ORGANISM	Chlamydomonas reinhardtii.		Organism Chlamydomonas reinhardtii.
COMMENT	Chlamydomonadaceae; Chlamydomonas; Chlorophyceae; volvocales; Chlamydomonadaceae; Chlamydomonas			Chlamydomonadaceae; Chlamydomonas
REFERENCE	AUTHORS	Grossman,A., Chang,C.-W., Davies,J., Hauser,C., Lefebvre P., McPermott,J.P., Shrager,J., Sillflow,C. and Stern,D.		Author Grossman,A., Chang,C.-W., Davies,J., Hauser,C., Lefebvre P., McPermott,J.P., Shrager,J., Sillflow,C. and Stern,D.
TITLE	Analyses of the Chlamydomonas reinhardtii Genome: A Model Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1031			Project Title Analyses of the Chlamydomonas reinhardtii Genome: A Model Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1031
JOURNAL	COMMENT	Unpublished (2001)		Unpublished (2001)
source	source	Contact: Charles Hauser		Contact: Charles Hauser
note="A one-year-old plant from the natural mangrove habitat in Pichavaram, Tamil Nadu, India was collected and brought to the M.S.Swaminathan Research Foundation. The whole plant was treated with the nutrient solution supplemented with 500 mM NaCl for 48 hours. Poly(A+)mRNA from the leaf tissue was purified and cDNA was prepared using Superscript Lambda System (Life Technologies, Cat. No. 19663-014). The cDNAs were size fractionated over Sizeselect 400 spin column (Amersham-Pharmacia Biotech Cat. No. 27-5105-01) and cloned in 5'Sal I - 3' Not I of pSPORT1 (Life Technologies Cat. No. 15383-011). The ESTs were sequenced from the 5' end using M13/pUC18 reverse primer in an Automated Sequencer (ABI310, Applied Biosystems) and submitted after editing to remove the vector and adapter sequences. BLAST search in BLASTX (Non-redundant) was carried out on 7th August, 2000 using default parameters and the results reported under putative identification/comment."		Source organism="Chlamydomonas reinhardtii" /strain="CC-1690 wild type mt+ 21gr" /db_xref="taxon:3055" /clone_id="C. reinhardtii CC-1690, Stress II (normalized)" /Lambda_Zap_II" /note="Vector: pBluescript II SK- Site_1: EcoRI; Site_2: XbaI; Stress condition II library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (NH4+ - containing) and shifted to TAP - NO3- (24hrs); H2 production conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant		
BASE COUNT	118 a	70 c	95 g	122 t
ORIGIN				



/note—"this sequence is part of a project of EST assembly resulting from the application of public contracts to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project."

		Query Match	Score	Length	DB	Match	
QY	274	ATGGTGAACTTACATCCCGTTTACATACCTGTTCCACCTTCCCCATGTAGAGATA	333	46.2;	DB 13;	Score 46.2 ;	
Ddb	570	GTCCTGAACCCAGGCCGACAAAATTGTTGATTCGGCTTCAACATCACATGTAGAGTC	629	44.4%;	Pred. No. 0.084;	Best Local Similarity 44.4%;	
QY	334	ATGGCAAAGTTCGGAAGACCCCTCTAAAGTTCAACTGGACGAAACHTGATATA	393	0;	Mismatches 233;	Matches 186; Conservative	
Ddb	630	GAGCTCAGTCATGGCAGTGTATCGAAGTTCAGACTGGCCGATTTTCCCTA	689	0;	Indels 0;	Gaps 0;	
QY	394	GACTTAGAAAGAAGTATTGAAATTAAATAGAGAAAGAAAAACCCCTTCGGGTPACTTGCT	453	4.38;	TTCGGGTTAAAGGTGACCTCATATCTACCTCTCTCAA	208	
Ddb	690	GATGTTGATCTCATTTGTCGAATGGTGAACAGCAAATGCCAAATGATATA	749	4.38;	ACAAATACCCGACCCGGAGCGAAAGAGTTAAAGCGATTTCGGCG 208	149	
QY	454	TACCCAACAACCCCACCGGAAACCTTTCCAGGGAAAATTGAGGAGATAAGAAC	513	4.38;	ACATCTACCGGACCCGGGACCCGGCCCTGGGATGAGGCTGGCTCATGC 293	234	
Ddb	750	TCCCCAAATAATCGATGGCTGATGAGGTTAAAGATGATCTGACATGACATGAC	809	4.38;	TTAAGGAAGAAAATTAGTTAGTCTTCGGTAAACGGTGGACTATCCCTCTCAA	268	
QY	514	AGGGTTTTCTGTTAATAGZGAAGCCTAATATCATTACT	557	4.38;	Db 294 TCCCGATGGAGAACCTGCTGGCTCCGGGACAGCTGATGCTCATGC 353	290	
Ddb	810	CTCCCCAAATACTCTAGCTGGTGAACCTTATTTGAGTTTC	853	4.38;	Db 269 TAGCTTATGGTGAACCTTTACACCGTTTACATACCTGTTACATCCCTGAGCTCCATGTAG 328	269	
RESULT 15	BM003279	662 bp	mRNA	linear	EST 25-OCT-2001	QY	
DEFINITION	lambda Zap II Chlamydomonas reinhardtii CC-1690, Stress II (normalized)					QY	
LOCUS	131108E09	15	Y1 C. reinhardtii CC-1690, Stress II (normalized)			QY	
ACCESSION	BM003279					QY	
VERSION	BM003279.1					QY	
KEYWORDS	EST.					QY	
SOURCE	Chlamydomonas reinhardtii					QY	

Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

Chlamydomonadaceae; Chlamydomonas.

REFERENCE METHODS Crosses 1 (bases 1 to 662) Crosses 3 (bases 1 to 662) Crosses 5 (bases 1 to 662)

*humors* Grossman, A., Davies, C., Shradar, J., Silflow, C., Lerebvre, B., Hauser, D., Starn, D.

TITLE : Analyses of the Chlamydomonas reinhardtii Genome: A Model Organism for Plant and Cell Biology

## Unicellular System for Analyzing Gene Function and Regulation in

Vascular Plants. Project: 1031

JOURNAL Unpublished (2001)

COMMENT Contact: Charles Hauser

DCMB Box 91000

Duke University Durham NC 27708-1000

Tel : 919 613 8159  
Fax : 919 613 8159  
E-mail : [info@vista.com](mailto:info@vista.com)

Fax: 919-613-8177

Email: chauser@duke.edu  
Location: Qualifiers

Search completed: April 15, 2003, 15:47:23  
Job time : 2164 secs

